7-16-2019

Multiple Identity Tracking and Motion Extrapolation

Ashley Buck
aeb1530@rit.edu

Department of Psychology, College of Liberal Arts

Rochester Institute of Technology

Multiple Identity Tracking and Motion Extrapolation

by

Ashley Buck

A Thesis in

Experimental Psychology

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

July 16, 2019

We approve the thesis of Ashley Buck:

Esa Rantanen, Ph.D. Date

Associate Professor, Department of Psychology

Faculty Adviser and Chair of the Thesis Committee

Eric E. Geiselman, M.A. Date

Senior Engineering Research Psychologist,

Air Force Research Laboratory,

Wright-Patterson Air Force Base

Reader

Paul R. Havig II, Ph.D. Date

Senior Engineering Research Psychologist,

Air Force Research Laboratory,

Wright-Patterson Air Force Base

Reader

Acknowledgements

I would like to thank my thesis committee members, Dr. Eric Geiselman and Dr. Paul R. Havig II, for their time and input into my thesis study. Their expertise have been invaluable in the development of this study. I would also like to thank Dr. Esa Rantanen for dedicating hours of time helping me and guiding me through the thesis development process. I am grateful for the immensely valuable feedback and advice all of you have provided me. I would like to thank Laura Kuljanova and Graham Home for their assistance in developing a study program and executing test sessions. Finally, I would like to thank my friends and family for their love and support along the way.

Abstract

Multiple Identity Tracking (MIT) is a paradigm in which individuals track the location and identity of moving objects in the environment. The first goal of the present study is to determine if individuals are able to extrapolate the position of moving objects and their identities while the objects are occluded. There is conflicting research on the source of a decline in tracking ability. Either the amount of time an object is occluded for, or the distance an object moved during an occlusion (i.e., displaced) could equate to a decrease in performance. The present study aimed to evaluate which variable (occlusion time or object displacement) is more detrimental to performance. The second goal of the present study aimed to address was whether individuals are able to complete a secondary task while tracking objects. The secondary task was timed with the goal of interfering with the maintenance rehearsal of objects. By doing so, the present study evaluated tracking ability through an "occlusion" that involves performing a task, as many realistic occlusions occur. Twenty-five participants tracked five moving objects with unique identities over 100 trials. Response time and number of objects checked were recorded. The results indicated that participants could keep track of the objects through an occlusion with 59% accuracy. There was a difference in response time performance between slow moving and fast moving objects when they were occluded for 2 seconds, but not 4 seconds. The results suggest that tracking multiple moving objects and their identities while performing to a secondary task during an occlusion is possible, without detrimental performance during a secondary task for most individuals. Additionally, we observed a task switching cost, with participants taking longer to find the first object compared to subsequent objects.

Contents

List of Figures

List of Tables

Introduction

Multiple Identity Tracking

Premise 1: Multiple Identity Tracking is a resource-consuming task.

Individuals are often tasked with tracking the location of multiple moving objects at once. For example, while driving, it is important to maintain an awareness of the locations of other cars and pedestrians in the vicinity to prevent getting into an accident. An additional example that occurs regularly for many people is object tracking while playing sports. The maintenance of the location of other players in the field is crucial to avoid colliding with another players. These tracking tasks, even ones that seem relatively routine or mundane on the surface, are challenging and demanding on our perceptual-attentional systems (Oksama & Hyönä, 2008). Humans have limited attentive resources and can only selectively attend to one region of the visual field at once (Z. W. Pylyshyn & Storm, 1988). Our environment is dynamic and can be unpredictable, which can make tracking tasks daunting and challenging (Oksama & Hyönä, 2008). The consequences of poor object tracking performance can be severe. Failure to accurately track moving objects at an optimal level puts oneself and others at risk, whether the consequences are as harmless as bumping into another person or as serious as permanent injury or a fatal accident.

Multiple Object Tracking. Multiple Object Tracking (MOT) is a paradigm first introduced by Z. W. Pylyshyn and Storm (1988) to study how individuals track multiple moving objects in a laboratory setting. A MOT paradigm typically involves continuously tracking a set of identical moving stimuli (e.g., a number of colored circles) within a computer simulation. An array of these items are identified as targets at the beginning of the trial while the rest of the objects serve as distractors. After a predetermined period of time, the stimuli are briefly occluded (i.e., hidden from view) behind a mask while the objects continue to move behind the mask. At the conclusion of a trial, all objects are unmasked and the participant is instructed to identify the location of a specific target as quickly as possible. This setup allows researchers to determine how individuals are able to keep track of moving objects under a diverse range of circumstances.

The original paradigm, developed by Z. W. Pylyshyn and Storm (1988), tasked participants with tracking moving crosses. In this scenario, ten crosses were present, and one to five of the crosses were identified as targets during each trial. The rest of the crosses served as distractors. Participants tracked the objects for a duration of time between seven to fifteen seconds as the objects randomly moved around a screen. After the set amount of time had passed, a square occluded one object. The participant was instructed to press a key if the occlusion was blocking a target object. MOT tasks such as these engage one's attention constantly, due to the dynamic, ever-changing nature of the situation (Meyerhoff, Papenmeier, & Huff, 2017).

Z. W. Pylyshyn and Storm (1988) hypothesized that to track multiple objects, preconceptual, bottom-up mechanisms automatically identify features of the visual environment, which are then used to form relational references between objects. More recent literature uses the term "preconceptual" rather than "preattentive" to avoid implying that attention does not play an active role in tracking situations (Z. Pylyshyn, 2004). It is presumed in the MOT paradigm that a spatial index is assigned to each object, which is referred to as a FINST (Z. W. Pylyshyn & Storm, 1988). FINSTs specify the approximate location of objects within the space the targets occupy without necessarily providing information about the object's identity (Oksama & Hyönä, 2008). A FINST is formed using the individuating features (i.e., the temporal continuity of an object's identity and the spatial relationships relationships between them) of each object to differentiate each object from the rest (Z. W. Pylyshyn & Storm, 1988). For example, an individual would be able to track two objects by observing that one object is "above" or "inside" another object, because the spatial relationship between objects helps differentiate each object and provides referential information (Z. W. Pylyshyn & Storm, 1988). Performance in this type of task heavily engages one's visual and informational processing capabilities due to the spatial and temporal relationship that allows each object to be differentiated, (Meyerhoff et al., 2017).

The formation of a FINST creates a binding that moves as all of the objects' features move across the retina. The tracking process using FINSTs occurs in parallel (i.e., all objects are tracked at the same time), but if objects are selected for processing in greater depth, objects may be processed further in a serial fashion (i.e., individually attending to one object at a time). This model suggests that the process of tracking object locations occurs partially in parallel, with a serial process occurring simultaneously as needed (Z. W. Pylyshyn & Storm, 1988).

The MOT paradigm has held relevance through decades and is the basis of current psychological research on multiple object tracking (Meyerhoff et al., 2017). Since the original work of Z. W. Pylyshyn and Storm (1988), MOT paradigms have been a popular area of focus in cognitive and perceptual research because it is widely believed that the attentional processes in MOT can be applied to everyday scenarios (Meyerhoff et al., 2017). As a result, researchers have developed paradigms that deviate from MOT to expand the application of object tracking research to scenarios that are not captured within the MOT paradigm. One of the most consequential developments from MOT is the MOMIT model.

The MOMIT Model. Realistic objects are usually not identical. For example, when you are driving a car, the cars in your vicinity are different colors, models, and have varying importance (emergency and construction vehicles may require one to perform a sudden action, which warrants using more caution around these vehicles). For some operational professionals, it is critical to keep track of multiple moving objects with unique identities (e.g., in air traffic control and military aviation). In these scenarios the MOT paradigm does not completely represent the true nature of the task because the objects being tracked are no longer identical (although they might be similar in appearance) and each object must be tracked as a uniquely identifiable object.

Multiple Identity Tracking (MIT) is a task which requires individuals to track and maintain the identity (i.e., identifying features) of multiple moving objects and the objects' locations concurrently (Oksama & Hyönä, 2008). MIT is more taxing on the cognitive and perceptual systems of individuals than tracking tasks which do not require the maintenance of identity information (i.e., MOT tracking). The Model of Multiple Identity Tracking

(MOMIT) was designed to capture an additional layer of complexity that occurs in tracking scenarios that was not accounted for in previous MOT tracking models; in certain scenarios, not only do the locations of individual objects need to be known, but the identity of each object must be tracked as well. The MOMIT model outlines the dynamic process of tracking the location of multiple moving objects, and binding object locations to their discrete identities. This is often referred to as the "what-where binding problem" (Oksama & Hyönä, 2008). Previous findings suggest that the identity component in tracking tasks is quite difficult and the ability of individuals to maintain location information is markedly higher than their ability to maintain information about object identity (Oksama & Hyönä, 2004; Horowitz et al., 2007).

The MOMIT model is based on five assumptions, also known as tenets (Oksama & Hyönä, 2008). They are: (1) To maintain bindings between each object's location and identity, the bindings between the object and identity must be effortfully refreshed in a serial fashion. To refresh the bindings, an individual must make continuous attentional shifts to each target and identify each target. (2) The ability to track objects is limited to a small number of objects with an average maintenance performance of about four objects at once. (3) Tracking ability improves when an individual tracks familiar objects due to an integration of long term memory representations in the formation of bindings. (4) Spatial indexes are formed through continuous attention switching and are stored in one's visual short term memory (VSTM). This index formation process results in an error between the object's location and the index that was formed when the object was last attended to. In other words, the representation of the location of objects in the VSTM quickly becomes outdated and falls out of alignment with the current location of the object only moments after the object has been attended to. (5) Finally, objects in the periphery can be attended to through a parallel process but not enough spatial information is indexed in this process to be able to differentiate distractors from targets (Oksama & Hyönä, 2008, 2016).

To track multiple moving objects and their identities, several components of memory have been implicated in the MIT process and are proposed to function together in the

MOMIT model. The model was formed to identify the memory stores and attentional processes that form the components of the semantic identity of objects, location information, maintenance of indexed location information, attention switching, and maintenance of the "what-where" bindings. The memory stores that have been implicated in the MOMIT process include the episodic buffer, VSTM, and long-term memory (Oksama & Hyönä, 2008).

Visual Short Term Memory (VSTM) is a limited store containing representations of visual input and the VSTM is a central component of the working memory model (Baddeley, 2000). Working memory consists of the central executive, the phonological loop, the visuo-spatial sketchpad, and the episodic buffer (Baddeley, 2000). The visuo-spatial sketchpad is used to process information in the VSTM. The VSTM is the memory store associated with the creation of spatial indexes (FINSTs) that compose the location-identity bindings of objects (Z. W. Pylyshyn & Storm, 1988). These indexes allow an individual to attend to other objects in the environment and maintain a general idea of the location of previously attended objects. The problem with the formation of indexes is that as they are created, they become outdated as objects move. This gives an individual a general idea about the location of an object, but the indexes do not move along with the object. This is referred to as the "VSTM coordinate error" by Oksama and Hyönä (2008). In accordance with FINST, it is presumed that the process of identifying location information operates in parallel but it deviates from the FINST model in that this only applies to objects in the periphery and does not include identity information (Z. W. Pylyshyn & Storm, 1988; Oksama & Hyönä, 2008, 2016). In other words, objects in the visual periphery can be tracked at the same time preattentively but identifying information cannot be tracked in this manner. Notably, only location information can be maintained in one's peripheral vision.

Cohen, Pinto, Howe, and Horowitz (2011) conducted studies directly comparing MOT and MIT performance, and found that location and identity information are both likely tracked using the VSTM. Oksama and Hyönä (2016) have researched the MOMIT model using eye tracking, and their findings support the theory that serial processing occurs when identity information is tracked, while location information alone can be tracked in parallel.

Long term memory is associated with the creation of bindings in the episodic buffer within the working memory. Tracking performance in previous MIT tasks has been superior in conditions where targets were familiar objects, indicating that LTM can play a role in reducing the difficulty of maintaining object identities. One's LTM interacts with the episodic buffer to increase the speed of the identity-location bindings' formation if the objects being processed are familiar to the individual. These bindings are more efficiently created and maintained than in cases where objects are not familiar (Oksama & Hyönä, 2008).

The episodic buffer has been presumed to be the temporary storage location of the location-identity bindings that are needed to perform in MIT tasks, and the episodic buffer is where bindings are created (Oksama & Hyönä, 2008, 2004; Baddeley, 2000). The information related to identity about objects is presumed to be processed separately from location information. The episodic buffer holds the what-where binding information in storage while objects cannot be attended to. The buffer can only maintain this information for a brief amount of time until objects can be attended to again, and the amount of information that can be stored here is limited (Oksama & Hyönä, 2008). In an experiment by Oksama and Hyönä (2004), it was demonstrated that when working memory capacity was tested against tracking ability, MIT performance decreased as working memory capacity was made less available.

The final processes in MOMIT are based on a higher order serial attentional process (Oksama & Hyönä, 2004). One's ability to track the identities of multiple objects depends on attending to each object serially. This specific attentional process requires a mechanism that determines which target to attend to next (Oksama & Hyönä, 2008). The decision making process requires strategic tracking rather than approaching the task using a random process, which would be both inefficient and inaccurate. It is necessary to use strategic processes to avoid losing the bindings of objects completely that may be missed with the latter type of strategy. Oksama and Hyönä (2008) have proposed that the ability to strategically attend to targets includes using each spatial index in the VSTM combined with the ability to process the location of other objects in the environment in parallel to choose the next target to attend to.

Fit of the MOMIT Model. Previous researchers have developed various parallel and serial processing models. The MOMIT model is developed based on a mix of these findings that best represent paradigms that involve object identities bound to the objects (Oksama & Hyönä, 2004). There are some discrepancies between MOT and MOMIT in how each model proposes one's attentional and perceptual systems function during a tracking task. MOT as a paradigm considers object tracking alone, without the inclusion of differing objects and their semantic properties (Z. W. Pylyshyn & Storm, 1988). The MOT theory proposes that the binding process occurs in parallel using the spatiotemporal properties of objects, which is a low-level, preattentive process (Z. W. Pylyshyn & Storm, 1988). The MOT model argues that the spatiotemporal location of spatial indexes move along with the objects, meaning there is no disparity between indexes that are not currently being attended, and the object's current location (Z. W. Pylyshyn & Storm, 1988). MOT is at odds with MOMIT in this regard. Logically, these proposed differences align with the differences between the function of each type of tracking task due to the added complexity and difficulty that identity information adds to the problem. MOMIT is the best fitting model for tasks involving the creation of identity-location bindings, as it has been demonstrated that this type of task is too demanding to be parallel in nature (Oksama & Hyönä, 2008).

Variables that Impact Tracking Performance. The MIT process is activated in a broad variety of environments and the properties of stimuli vary as well (Oksama & Hyönä, 2008). Depending on the scenario, the number of objects-to-be-tracked, physical features of the objects, speed, direction, and entropy (entropy meaning the projected uncertainty of the trajectory of an object, considering the object's velocity, and both the frequency and magnitude of direction) might vary (Oksama & Hyönä, 2008; Hope, Rantanen, & Oksama, 2010; Keane & Pylyshyn, 2006). The number of objects that function as distractors and the types of object occlusions can vary, which can all contribute to poor tracking performance

(Oksama & Hyönä, 2008; Hope et al., 2010; Keane & Pylyshyn, 2006). Oksama and Hyönä (2004) found that individual differences play a significant role in performance in tracking tasks as well. The ability for individuals to track multiple objects with identity information was compared to tracking tasks that do not consider identity information, and it was found that performance significantly decreased when identity information must be tracked in addition to location information (Cohen et al., 2011). Researchers have examined the relationships that exist between these variables and human performance in tracking tasks, finding which aspects lead to a decline in performance and why.

There are some variables that are present in the MOMIT paradigm that have been examined in depth. A robust finding is that as the number of objects present in the tracking task increases, performance in the task decreases (Oksama & Hyönä, 2004; Z. W. Pylyshyn & Storm, 1988). Generally, 4-5 targets are able to be maintained successfully in the VSTM, and this number can be maintained in a field of up to 8-10 objects if distractors are used (Oksama & Hyönä, 2008; Z. W. Pylyshyn & Storm, 1988). When comparing performance directly between MIT and MOT tasks, location accuracy was impacted by the additional load of tracking identity information, and individuals are able to distribute their mental resources flexibly to maintain location and identity information (Cohen et al., 2011). There are also interactions between these variables that have been found to decrease performance. For example, as object speed and the number of objects increases, performance deteriorates at a steeper incline. Additionally, the number of objects and familiarity of objects also reduce performance at a steeper incline (Oksama & Hyönä, 2008). This knowledge can be applied to scenarios where the number of objects present in a known tracking paradigm could be limited. However, there are some aspects of object tracking that are not as well understood, or present conflicting findings in different studies.

Object Occlusions and Tracking Performance. Occlusions occur from eye saccades, physical head movement, or from an object that blocks the view of targets (Keane & Pylyshyn, 2006). We encounter such occlusions everyday; objects disappear from view behind trees and buildings, while clouds and poor weather can limit visibility. In operational settings, a lost signal can result in the objects disappearing from a screen, or an operator turning their attention elsewhere (e.g., briefly looking between multiple displays). It is unlikely in realistic scenarios that occlusions can always be prevented, yet there is little known about the mechanics of how these occlusions can impact performance. While tracking multiple moving objects, one's tracking ability might be limited due to the distance an object traveled (displacement) while it was occluded or the amount of time an object was occluded. Having an understanding of the limitations of our tracking abilities in scenarios where objects are occluded from shifting attention elsewhere is important for improving the safety and performance of those who use MIT in operational settings.

Motion extrapolation is an ability that has been questioned and examined by MOT researchers with conflicting findings. Motion extrapolation is the premise that, "if an observer can project with reasonable certainty where each object is headed, he or she should be able to use those projections in the interpretation of noisy measurements of future positions" (Zhong, Ma, Wilson, Liu, & Flombaum, 2014, p. 2). It has been found through several research paradigms that motion extrapolation is a task that does not yield high performance. The complexity of the displacement problem is amplified by uncertainty in the environment, whether it is due to exogenous or endogenous reasons (Moray, 1984). An exogenous reason would result from an external characteristic of the environment, while an endogenous reason would occur from an internal psychological phenomenon, such as fatigue. To counter uncertainty in a tracking task, the user within a system must counter the likelihood of uncertainty through maintenance of bindings (Moray, 1984; Oksama & Hyönä, 2008).

Keane and Pylyshyn (2006) investigated a known phenomenon in visual tracking, the prediction hypothesis, in which objects that momentarily disappear are perceived to continue to persist as smooth linear projectiles, as long as they maintain the same velocity and trajectory that they had before the interruption occurred. Their initial 2006 study examined the relationship between different durations of occlusions and displacement in a MOT task. The findings suggest that tracking performance is best when the objects do not move far from the position they were occluded in. In this study, participants were able to track objects that were occluded for up to 900 ms with 80-90% accuracy, but performance dropped off as a function of the magnitude of the object displacement. It can be concluded from these findings that prediction is generally not employed in MOT paradigms, meaning that as individuals are tracking objects, they are not able to predict the object's future location . Keane and Pylyshyn (2006) also found that FINSTs were persistent through occlusions. Fencsik, Klieger, and Horowitz (2007) also found similar results to Keane and Pylyshyn (2006) . Based on the ability to track objects through long occlusions, it can be concluded that iconic memory is not used in tracking multiple moving objects. When objects disappear, there is evidence of the "persistence of position" that maintains the location of objects (Keane & Pylyshyn, 2006).

When Franconeri, Pylyshyn, and Scholl (2012) examined the relationship between motion extrapolation through occlusions, they found that performance was best when the objects had moved a short distance from where the object occluded, despite the amount of time it took to reach that point. In similar research studies, performance has been best in paradigms where objects remain visible throughout the task, rather than when a masking condition is used (Cohen et al., 2011; Z. Pylyshyn, 2004; Oksama & Hyönä, 2004). Furthermore, the particular path that the objects are moving on did not seem to impact performance in these studies, as long as the motion was predictable in that it was not erratic (Howe & Holcombe, 2012). A study by Hope et al. (2010) found that tracking ability increased as an inverse function of object entropy (i.e., performance decreased as the movement of objects was more predictable). These studies have provided strong evidence that individuals struggle to extrapolate the trajectories of objects during tracking tasks.

Contrary to these findings, other research studies have found evidence in support of motion extrapolation. One study concluded that participants encode the directional information of moving objects because performance had been negatively impacted by patterns on moving objects that moved counter to the movement of objects before they were occluded (Clair, Huff, & Seiffert, 2010). Another study found that participants are able to identify the direction that objects were moving in a tracking task before an occlusion (Shooner, Tripathy, Bedell, & Öğmen, 2010). Luu and Howe (2015). The mixture conflicting evidence about the nature of object extrapolation and displacement in previous research has not thoroughly been resolved.

A computational examination of a MOT motion extrapolation paradigm was performed by Zhong et al. (2014). These researchers determined that poor performance in previous studies possibly stems from inaccurate extrapolation, rather than a complete inability to extrapolate motion. Furthermore, they believe has not been captured because of the way the results had been measured previously. Their findings demonstrated that tracking multiple targets in noisier environments led to conservative estimations of where objects were located after interruptions and occlusions (which is indicative of maintaining Level 3 SA). This may explain various findings in the literature, but unfortunately suggests that previous conflicts arose from characteristics of how these studies are conducted and not participants' tracking ability. This study also provides a balanced view that may explain the discrepancies between findings that support or argue against the abilities of individuals to extrapolate motion.

These examinations of object occlusion and speed are limited in MIT research. Hope et al. (2010) concluded that a possible source of their unexpected finding was that objects with larger entropy moved a shorter distance from their original position during occlusion and were therefore easier to reacquire after the occlusion. This study continues this research and systematically examines the relationship between object displacement and object occlusion time. The findings in MOT research appear to have stronger support of an inability to project object trajectories over long displacements, but there is not a definitive consensus in MIT research.

Situation Awareness

Premise 2: Multiple Identity Tracking is a paradigm much like Situation Awareness.

A model of Situation Awareness. Situation Awareness (SA) is a theoretical cognitive construct proposed by Endsley (1995b) that can be used to help understand the attentional processes in the present study's experimental paradigm. SA is a widely used theory in the field of engineering psychology in system and interface design. SA is defined as, "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1987, p.65). SA is characterized in three levels, with Level 1 being the most basic perceptual conception of the environment, and Level 3 being a highest level understanding of a system. As an operator advanced from Level 1 to Level 3, their capacity to operate a system successfully increases.

Level 1 SA in Endsley's model is defined as the ability to perceive the environment and features of its elements, such as color, brightness, sound, and object size (Endsley, 1995b). Level 1 SA is the basis of which an operator's understanding of the environment is built upon, and it is critical in task performance. As the basis of more advanced levels of SA, it is critical that operators perceive the necessary elements of the environment to successfully accomplish a goal. It is also required that these elements are perceived accurately. For example, it is critical that a fighter pilot is able to gauge elements of the environment such as altitude, heading, targets, and the status of the system. If a pilot misreads his or her heading, or if the display is damaged, the pilot would be in immediate danger. Luck may help avoid a disaster, but the very basis of the pilot's understanding of the environment and situation would be inaccurate. Therefore, any decisions and actions based on that heading would be formed on incorrect information about the environment, which could have costly and deadly consequences (Endsley, 1995b).

Level 2 SA is acquired once an operator has an understanding and comprehension of the environment and situation (Endsley, 1995b). Based on the goals of the individual operating within the system, specific elements of Level 1 need to be integrated into the operator's cognition of the scenario. Additionally, the elements that are relevant to the current goal need to be understood well enough to accomplish these goals. The environmental elements that are necessary to successfully achieve Level 2 SA are unique to each scenario, and it is dependent on the operator to know and integrate the correct elements into their perception of the environment for successful performance.

The most advanced stage, Level 3 SA, involves a projection of the future status of the environment and its elements. Based on one's current understanding and knowledge of the system, decisions can be made about one's best course of action to achieve goals or avoid danger. Having Level 3 SA is an ideal state of an operator to make informed decisions and perform well in dynamic situations (Endsley, 1995b). An operator performing at a lower level of SA would be at risk of making mistakes based on a misunderstanding or inability to understand how their decisions will impact the future state of the system.

The elements of the system are determined by the "spatial, temporal, or functional relationships of elements to goals" (Endsley, 1995b, p.38). Due to the dynamic nature of most environments, and the level of knowledge and experience needed to understand a system well enough to make informed decisions, time is essential in the development of Level 3 SA. Therefore SA is considered to be highly temporal in nature. Spatial information also plays a role in the development of Level 3 SA (Endsley, 1995b). Depending upon the operator's task, the interactions and physical location of other objects in space play a role in one's understanding of a situation.

Relationship between MOMIT and SA. Based on the description of SA above, it is clear that there is overlap between the functionality of the processes of MOMIT (MIT) and SA. In the process of tracking objects, individuals must be able to extrapolate motion, or in other words, make predictions about the future position of moving objects (Oinonen, Oksama, Rantanen, & Hyönä, 2009). The present study proposes that to make predictions about the future location of a target object, an individual must maintain Level 3 SA. In an MIT tracking paradigm, participants must continue to track the location and identity of an object through a mask, however brief the occlusion. If an individual has not achieved Level

3 SA, then making predictions about the future location of targets once they are unmasked would not be possible.

It has been demonstrated that the identity tracking aspect of MOMIT involves direct attentional process and involves an integration of working memory with LTM (Oksama & Hyönä, 2008, 2016), meaning that this process is more advanced than a low-level parallel perception of objects. Each object must be serially processed by integrating its FINST with information about other objects in the environment from the VSTM (Oksama & Hyönä, 2008). In addition to the mismatch between the location of a FINST and the object's current position, there is a processing lag as well (Howard, Masom, & Holcombe, 2011). Therefore, to be able to correctly track objects, it is important to be able to make predictions about the future location of an object. It is likely that one's ability to extrapolate motion occurs due to a neural compensation for the lag in processing at the current location of a moving object (Howard et al., 2011).

Historically, SA has been evaluated using subjective reports and objective performance-based measures (Endsley, 1995a, 1987). SAGAT (Situation Awareness Global Assessment Technique), developed by Endsley and Garland (2000), freezes a simulation at a random interval and queries the participant about the simulated situation at that point in time. This dynamic is similar to an object tracking paradigm, in which the objects are frozen and the participant is tasked with identifying the location of an object. By measuring reaction time and error rates, it can be assessed if a participant was able to maintain a degree of awareness of the situation during the tracking scenario. Future developments in the MOMIT paradigm could further link measures of SA to measures of tracking ability.

Applied Context

Premise 3: The MIT paradigm is readily applicable to operational settings.

The development of the MOMIT paradigm emerged as the realization came about that MOT is not always applicable to realistic scenarios (Nalbandian & Rantanen, 2015). The application of MIT research to operational settings is invaluable, but not all research can be generalized to all operational scenarios. Studies that emphasize realistic paradigms and characterizations of objects are bound to provide information that is useful for understanding task performance that we might expect to see from operators in the real world. Several aspects of previous studies, particularly the use of occlusions and target speeds, could use further adaptation to be more relevant to fighter pilots and air traffic controllers (ATCos).

Previous studies that have examined the prediction hypothesis and object tracking generally use occlusions below 1 second, with 900 ms being the longest duration to our knowledge. Participants are able to maintain 90% accuracy in their performance through long occlusions such as these (Keane & Pylyshyn, 2006). Other studies also use a range of target speeds as fast as 20.5 degrees visual angle/s (Cohen et al., 2011). While there are scenarios where objects move quickly and are briefly occluded, such as sporting events, the use of short occlusions and fast moving targets is not applicable to tasks related to the work of military pilots and ATCos. For example, Nalbandian and Rantanen (2015) examined MIT in a scenario that replicates the tasks of ATCos. The study was designed to use objects that replicated the call signs of actual aircraft, with object movements and trajectories that resemble the movements that objects would make across a typical ATC display. The performance of ATCos varies greatly from MOT and MOMIT due to the slow speed that the objects appear to move on an actual radar screen, and the large number of objects that ATCos successfully track (up to 20-30 at a time). To change the paradigm's applicability, the present study extended the duration of masks while using objects that generally move slower than in previous research studies to evaluate the question of object displacement versus object occlusion time. The use of a longer mask can be representative of the work of individuals in operational settings because it would reflect tracking performance when an operator must look away from a task for a certain amount of time. The use of long mask durations has the potential to yield high success rates in performance when paired with slow moving objects, and the relationship between these two variables could use elucidation.

The present study used several additional adaptations that are not novel, but defer from most MIT paradigms. If an operator, athlete, or any individual were tasked with tracking moving objects, it would not be assumed that some objects in the environment can be ignored. It is probable that some objects are less significant to an individual than others. A soccer player in a forward position might not actively seek out the players on their own team who are in defensive positions or the goalie, but this does not dismiss the need to be able to differentiate those players from the rest and have a sense of their locations. Furthermore, the player would need to identify the individual to determine if they are of importance to the task at hand. The use of distractor objects might be relevant in certain scenarios, but there are scenarios where all of the objects present are of equal importance, such as in an Air Traffic Control scenario. As such, the present study excludes distractors and all objects used in the study are potential targets. As highlighted in previous studies, each object must be attended to in a serial fashion to be able to form an identity-location binding, while location alone can be tracked in parallel (Oksama & Hyönä, 2008; Z. W. Pylyshyn & Storm, 1988). Neither process supports the idea that only certain objects are attended to in the visual field. It is only once an object is identified that it is assigned importance. Therefore, the use of distractors can be limiting depending on the application of the tracking task.

Object occlusions in operational settings. An aspect of tracking paradigms involves the use of a mask, which occludes the objects being tracked for a designated amount of time (usually milliseconds) (Oksama & Hyönä, 2004; Keane & Pylyshyn, 2006). Masking is a known critical component of the MIT paradigm (Oksama & Hyönä, 2004). Most studies make use of what is called a blank screen interstimulus interval (ISI)(Keane & Pylyshyn, 2006). This type of mask is a blank screen which occludes all objects. Objects generally continue to move behind the mask, with the mask varying in duration, and objects may be fully or partially occluded behind the mask. The use of the ISI has been relevant to research on the prediction hypothesis. To understand the work of ATCos and military pilots, it must be considered that these individuals are tasked with tracking multiple moving objects and must perform other tasks at certain points. The use of an ISI alone, therefore, is not a good predictor of tracking performance when the occlusion occurs from the need to perform a secondary task.

Not much research has been conducted on the use of integrating a secondary task into a tracking study. To our knowledge, research has not been conducted where a secondary task has been provided on a mask. However, several studies have been performed where a secondary visual search task is performed simultaneously with the tracking task. These studies suggest that the use of a secondary task rather than an ISI is not only possible, but would still yield high performance.

Several studies have found that tracking can persist while one must simultaneously search through the visual environment to perform a visual search task (Alvarez, Horowitz, Arsenio, DiMase, & Wolfe, 2005; Horowitz, Birnkrant, Fencsik, Tran, & Wolfe, 2006). Alvarez et al. (2005) performed several simultaneous visual search and MOT experiments. They found that tracking performance was dependent on the visual attention resources that were required to perform each task. They concluded that there is likely a limited-capacity spatial memory that can be used to recover the trajectory and location of tracked objects once attention has been diverted briefly. Their results suggested that interruptions to tracking did not lead to a decrease in performance.

In an experiment by Horowitz et al. (2006) it was found that performance in the tracking task and visual search task was high, even with a gap of 300-500 ms where all objects were occluded. It was proposed that there are different mechanisms available that enable visual search and tracking to occur simultaneously, referred to as the switching hypothesis. The results suggest the presence of a flexible attentional system that allows individuals to switch attention between two demanding tasks.

Relevance to operational demands. MIT paradigms are highly applicable to the operational settings of fighter pilots. The F-35 Joint Strike Fighter is a tactical jet used within the Navy, Air Force, and Marine Corps (Gertler, 2012). Pilots who fly this jet are fitted with a helmet mounted display (HMD). The HMD features information about the status of the jet, such as altitude and speed, and is directly in the line of sight of a pilot. Pilots also have access to a horizontal situation display (HSD) that features information of other aircraft in the vicinity. The HSD is specifically a heads-down display (HDD), meaning the pilot must look down at the display. Elements of the HMD can be presented on the HSD. The pilot must maintain information between the different displays and the dynamic environment they are flying in. To view information on the HMD, the text must be foveated, meaning the application of a simultaneous task is not an appropriate representation of this display system. The nature of these displays creates an environment where long occlusions are inevitable, both in terms of angular display and distance. The ability to track the location and identity information of each aircraft in the surrounding environment is crucial to fighter pilots.

Military pilots can encounter object occlusions for an array of other reasons, too. Objects in the visual field can be occluded behind clouds or exit the visual field of the jet. The pilot may have to switch to a separate task that occupies their attention for a brief period of time, such as changing a communication channel to a different frequency. The separation of information based on angular distance between the HUD and HSD displays and the amount of time needed to switch between these displays could realistically take several seconds, rather than milliseconds. There is also an accommodation change that is necessary when switching between viewing the HMD and HSD, because the HMD is collimated at optical infinity, while the HSD has a shorter focal length.

Searching for information on the HMD is aided by maintenance of the identity-location bindings of aircraft in the area. Rapid acquisition of targets of interest is critical given the gravity of situations this aircraft is designed to be used in. Should the pilot forget the identity-location bindings, acquisition of a target will become a visual search task. Performance in this scenario goes beyond simply looking away from the MIT task environment (hence the lack of application of a plain ISI to this scenario). The HMD and environment are more representative of task switching scenarios, where the HSD would represent an MIT task and information on the HMD a mask occluding the HSD.

The use of a task switching task in place of a blank mask adds a realistic element to the current literature in MIT. A realistic scenario where objects are occluded could often result from having to engage in an alternative task. In a scenario where the environment is occluded due to an object blocking the view of the object, the operator would likely be able to engage in rehearsal to avoid losing the position of objects that must be tracked. In a scenario where the operator must look away because the operator needs to engage in a secondary task, the ability to engage in rehearsal would be lost. Therefore, the external validity of tasks in which a blank mask is used is limited to scenarios where individuals would be able to track objects without more than a visual interruption.

Purpose of the Research

The present study addresses several questions related to multiple identity tracking performance. The first is to identify the source of poor performance in tracking multiple moving objects while maintaining identity-location bindings. This study will expand on research investigating the relationship between object displacement and masking on tracking performance. Isolating and identifying the aspects of tracking multiple objects that are most detrimental to object tracking performance will help designers, engineers, and researchers identify solutions that would result in improved performance and a lighter cognitive load during tracking tasks. No study to our knowledge has examined if the displacement of an object or the duration of time that passes during an occlusion of an object has a greater impact on tracking performance in a multiple identity tracking task. Previous research studies on MOT have resulted in conflicting conclusions, making it unclear how the relationship between occlusion time and distance functions in any tracking task. Additionally, the mechanisms of MOT and MIT are presumed to differ, meaning the findings in MOT research are not necessarily applicable to the maintenance of identity information in tracking tasks.

This study is also an examination of tracking abilities with the goal of understanding how object tracking is performed in realistic operational scenarios where operators are tasked with tracking objects through extended occlusions to perform a secondary task. The need to look away from a screen to perform a secondary task is unavoidable in operational settings. Research has tackled how individuals are able to perform these tasks concurrently with object tracking (and participants perform well in these scenarios). The present study is novel in how a secondary task is used as a means of interference in identity-location binding rehearsal by presenting a secondary scenario as a mask. As a result of including such a task, this study uses occlusion times that are longer in duration than those in previous studies, with relatively slower object speeds. Previous studies suggest that this task is possible, but it has not been testing in a research setting before.

Hypotheses. Four combinations of object speeds and occlusion times were examined. The conditions were as follows; a fast moving object with short occlusion, slow moving object with a short occlusion, a fast moving object with long occlusion, and a slow moving object with a long occlusion. The use of a slow moving object with a short occlusion and a fast moving object with long occlusion created two conditions that were directly comparable, as they both had equal displacements. The other conditions created scenarios that were hypothetically easier or more difficult than the comparable conditions. It was hypothesized that object displacement would result in poor performance (higher error rates) in identifying targets compared to the duration of the mask.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \tag{1}$$

$$H_A : \mu_i \neq \mu_j \text{ for at least one } i \neq j \tag{2}$$

In a realistic scenario, this would mean that if a pilot lost track of a moving object when they foveate to the HMD, it would likely be due to the objects on the HSD moving far from the point they were last attended to, rather than the pilot spending more time foveating on the HMD. By comparing both Displacement B conditions, it is likely that if occlusion has a stronger effect, performance in the bottom left quadrant (Table. 2) will result in poorer performance than in the top right quadrant.

Table 1. *Object movements in experimental conditions.*
Displacements during each occlusion measured in degrees VA/s.

	Short Occlusion (2s)	Long Occlusion (4s)
Slow Speed (11.1 mm/s)	Displacement = 2.54	Displacement = 5.08
Fast Speed (22.2 mm/s)	Displacement = 5.08	Displacement = 10.08

Ethical Considerations

The confidentiality of all participants, whether or not they completed the study, was maintained to the fullest extent during data collection. To maintain the confidentiality of participants during data collection, participants were assigned an ID number that was used on data collection forms. The only personally identifiable information was the participants name and signature, which was collected on the informed consent form and in an electronic document where names were paired with an ID number in the event a participant must be identified for the researcher's purposes. The electronic document is the only location where the name of the participant and the link to the participant's ID number is located. This information is only accessible by the primary researcher, research advisor, and research assistant. Informed consent forms are kept securely in a locked research lab, while the electronic form documenting name and ID number pairings requires a security code to access the document. In the event information must be disposed of, it will be shredded and deleted to the fullest extent so identifying information cannot be recovered. The data collection process took participants 45 to 60 minutes to complete. Students had the opportunity to sign up to participate at a variety of times over several weeks, which did not require them to miss class to participate.

There were minimal foreseeable psychological, physical, social, or legal risks or benefits to participation. The informed consent process occurred between each participant and the researcher to prevent any harm to participants feeling coerced to participate. Participants were provided with information about the nature of the experiment and their rights as a participant. Deception was not used in this study. The purpose of this study is to test human performance capacity, and deception would not improve or cause a decline in performance.

Participants were informed that it would be possible to experience feelings of frustration during the study if they find the task to be difficult. Participants were reminded that the task was intended to be difficult, and is not related to IQ or their individual capacities in any way during the training phase of the study and the debrief session. Participants were compensated through course extra credit or entry into a raffle to win a $50 gift card.

Method

Participants

Twenty-five participants were recruited from Rochester Institute of Technology's graduate and undergraduate student population. The sample size was chosen based on $\alpha = 0.05$, $1 - \beta = 0.8$, and a small effect size ($\eta_p^2 = 0.02$). Two participants were removed from the study due to performance below thresholds set for object tracking performance and secondary task performance ($N = 2$).

The sample consisted of female ($N = 11$) and male participants ($N = 13$) of varying races (White, Asian, Biracial, Black, and Hispanic). Both participants that were removed were female bringing the total count to nine females ($N = 9$). The average age of the participants was 22.4 years old (range 19–29 years). Participants came from a mix of mostly technical and some humanities based fields of study, such as Computer Science, Industrial Engineering, Psychology, and Advertising and PR. All participants were evaluated based on threshold performance, so participants would be removed if they performed below a baseline level. All participants reported having normal or corrected-to-normal vision ($N = 24$). All procedures were approved by Rochester Institute of Technology's Institutional Review Board (IRB).

Apparatus and Stimuli

To run experimental trials, a program was created using Javascript and run in Java Runtime Environment which allowed the researchers to manipulate the experimental and masking conditions. The study was conducted in a quiet lab using a computer with a 22.5 inch monitor. Participants interacted with the program using a standard mouse. The background image was a satellite screenshot of a forested area. The features of the image were blurred using a blurring effect in Adobe Photoshop to remove any identifying features from the image that could be used strategically as reference points in the task (Fig. 1). An image was chosen rather than a solid background to add face validity and to hold participant interest in the task. All objects within the background and 10% of the screen surrounding

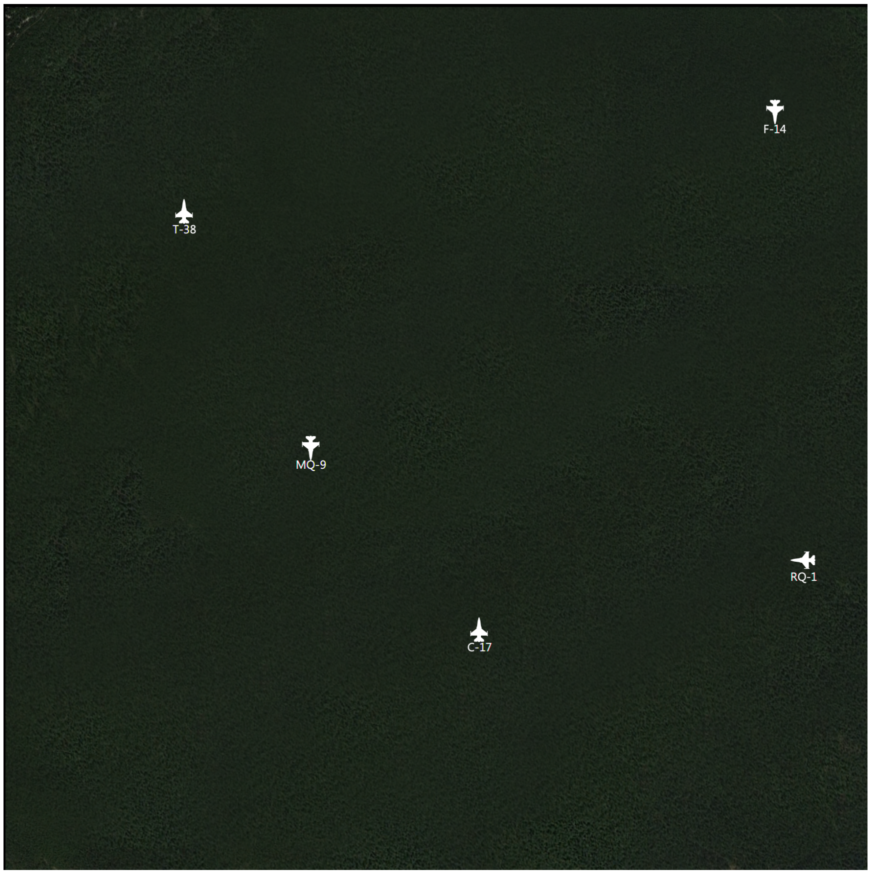


Figure 1. An image of the experimental paradigm.
A simple forested satellite image was chosen as the background image, with each object and a call sign identifier to be paired with the object as its identity.

the background are logged within the program, allowing for error beyond the perimeter of the background image to be recorded. The frame rate was set to 60 fps.

Objects were generated using icons in the font collection "Font Awesome." Each of the five objects was a small aircraft silhouette set to a font size 20 (which appeared to be 1 cm in size on the monitor). The font size was set to 12 for all object identity labels, and all objects and labels were white to ensure enough contrast between the background image and objects so visibility does not confound the results (Fig. 1). The font size allowed for readability but required the participant to foveate to the text to be able to read it clearly (Saha, Samanta, Sarcar, & Sharma, 2012). In addition to each object, a 3 digit call sign (an alphanumeric mix of either 2 letters followed by a number, or 1 letter followed by two numbers) was visible below each object to serve as its label and identity. For example, “RQ-1” or “C-17” was paired with each object (Fig. 1). These labels are based on the designators of real types of planes to add face validity for participants. A click radius was set to 10% of the total screen size, so participants only had to click on the relative location of the object. Accuracy was theoretically better reflected using this method due to the small size of the objects, and the potential for participants to miss the object slightly when attempting to click on it.

To determine the speed of an object on a plan view display (PVD), display scale, display size, and viewing distance needed to be taken into account to calculate visual angle per second. The average viewing distance was determined to be 500 mm from the screen, which was a close but comfortable viewing distance from the screen (Nalbandian & Rantanen, 2015). A standard 22.5-inch monitor (59.45 degrees VA across) was used as the display. Objects generated at one of five designated waypoints, with each object generating at its own waypoint. Objects moved in predetermined trajectories between waypoints in rectangular overlapping trajectories.

To determine the mask duration, two times were chosen that would allow participants to engage in a task that is provided to them on the mask. Two and four second masks were determined based on the amount of time that participants could complete a given task and remain occupied for the duration of the mask. Two seconds was determined to be sufficient based on pilot testing, then this value was doubled to create a long occlusion time. Object speeds were chosen and adjusted based on the duration of each mask (either 2 or 4 seconds) and the need to create equal displacement conditions between a slow speed and fast occlusion and a fast speed with a long occlusion.

To create comparable conditions between an object that moved slowly with a long occlusion and an object that moved quickly with a short occlusion, the objects needed to have equal displacements. The object speeds are not novel based on the fact that they move slowly, because other studies have used similarly slow moving objects, but the speeds

Table 2. *Values of object speeds that were used in this experiment.*
These calculations are based on a 22.5 inch PVD at a 500 mm viewing distance (59.45 degrees VA).

Condition	Occl. time (s)	Obj. spd. (mm/s)	VA/s	Displ. (VA)
Slow Speed, Short Occlusion	2	11.1	1.27	2.55
Slow Speed, Long Occlusion	4	11.1	1.27	5.08
Fast Speed, Short Occlusion	2	22.2	2.54	5.08
Fast Speed, Long Occlusion	4	22.2	2.52	10.08

of the objects in the present study are at the lowest end of the ranges of speeds typically used (Keane & Pylyshyn, 2006). The use of slow speeds with long occlusions resulted in displacements that are typical in object tracking research (Keane & Pylyshyn, 2006). The final object speeds can be viewed in Table 2.

Degrees of Visual Angle was calculated using the equation

$$VA = arctan(\text{speed})/500\text{ mm} \tag{3}$$

and displacement was calculated using the equation

$$D = \text{occlusion time } \times DegVA \tag{4}$$

Research has demonstrated that that the familiarity of objects is related to performance. As familiarity increases, error rate decreases (Oksama & Hyönä, 2008). The identities chosen in this study were kept constant and paired with simple white triangles as objects to eliminate the possibility that the familiarity of certain objects over others may have an effect on performance. Additionally, the identification of these objects was dependent upon the call sign, rather than the object icon as a means to reduce variability.

Previous research has indicated that tracking performance deteriorates as a function of the number of objects in the trial, with about 4 objects being maintained in the episodic buffer on average, and 4-5 objects being maintained successfully out of a total of 8-10 objects when distractors were used (Oksama & Hyönä, 2008; Z. W. Pylyshyn & Storm, 1988). Based on these findings, it was determined that the ideal target number should be 5 total objects at any given time, to avoid an unintended floor or ceiling effect based on the number of objects present. Object entropy was held constant for all objects to prevent an interaction between uncertainty in movement and the outcomes in performance, as Hope et al. (2010) had found a marginal effect of entropy on performance.

Design

The independent variables in the present study include object speed and mask duration, which both equate to object displacement. The faster an object is moving, the more it will be displaced in a certain amount of time, which is why manipulating object speed will provide insight on the effect of object displacement on tracking ability at constant occlusion time. Different durations of the mask and different object speeds allowed for study of the effect of occlusion duration while keeping displacement constant.

The dependent variables of interest were response time, identity check counts, and correct responses to the secondary task. Response time has been used previously as a measure of both situational awareness and multiple identity tracking (Oinonen et al., 2009; Endsley, 1995a; Hope et al., 2010). During each trial, if the participant was able to track the location of an object and maintain level three situational awareness, the amount of time it should take to locate the object should be less than that of an individual who did not maintain the binding and level 3 SA. Response time was measured within the program that was used to run trials with participants. The identities of potential targets remained masked when the secondary task mask was lifted (Fig. 2).

If the participant was unable to identify the correct target, the participant could reveal the label of the object identities by hovering over the object with their mouse ("mouse-over", which will be referred to as identity checks). The label mask was used to help differentiate between participants who had completely lost track of the objects and participants who happened to look to the correct object first or had a general idea of where the target was located. If participants are able to maintain a general idea of where objects are located, they

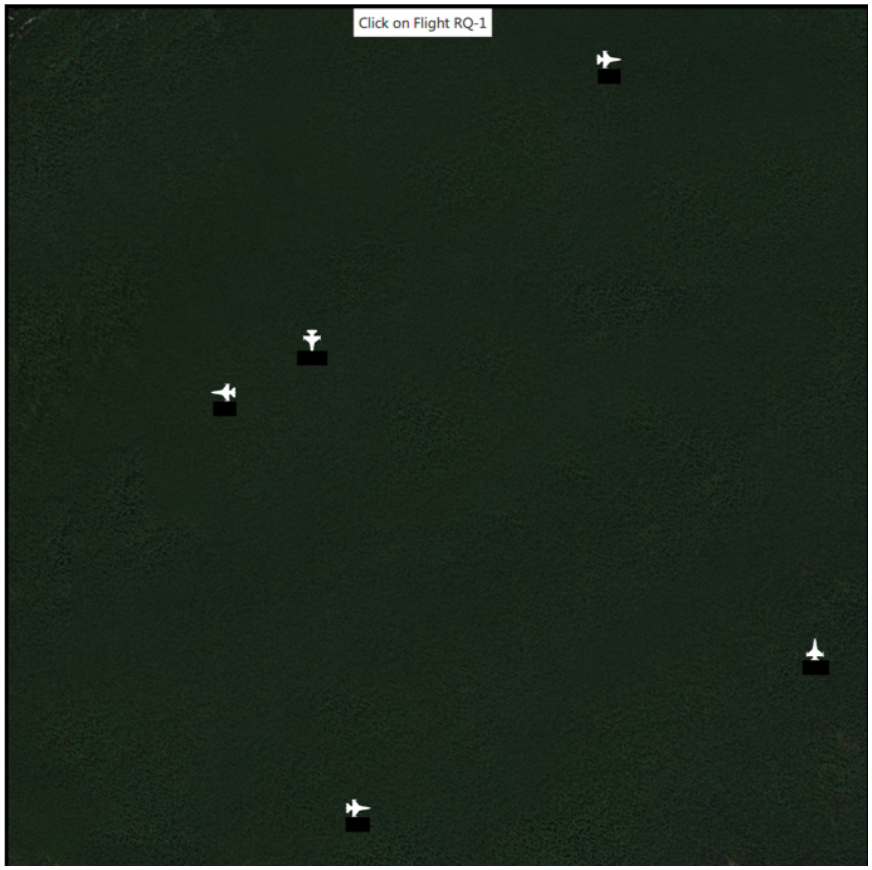


Figure 2. An image of the task once the mask was lifted.

should be able to identify the correct object without revealing more labels than the label of the target object. The program revealed the label if the mouse hovered over the object in any capacity, meaning an object could not be clicked on without revealing the identity. Hence 1 identity check per query represents perfect performance. If the participants were guessing, they would likely have to check multiple object identities before finding the correct object. With 5 objects, chance performance was 20%.

As a participant, it would be advantageous to ignore the secondary task and rehearse the object locations. To determine if participants were actively engaging in both tasks, the present study also accounted for performance in the secondary task. Participants were removed if they performed at less than chance (50%) in the secondary task appearing on the mask screen.

A factorial 2 × 2 within-subjects design was chosen, with 2 levels of object speed (slow and fast) and 2 levels of object occlusion task time (short and long). Trial orders were randomized to assign an order of appearance per trial for the four potential experimental conditions. Participants were randomly assigned to take the experimental trials in one of two randomized orders. Random assignment was used to prevent bias in grouping participants, and to prevent an order effect from confounding the results. The total number of trials per participant was 115 (15 practice trials and 100 experimental trials), which was determined to be a sufficient number of trials to reduce standard error per trial.

Procedure

Informed consent. To begin the experiment, participants were given informed consent forms. Participants were allowed to take as much time as they wanted to review the informed consent form with the researcher and ask questions. The researcher and research assistant used a script in this process to ensure all necessary requirements were fulfilled, and each item was addressed before the experiment began. The researcher reviewed the forms in depth with each participant, and gave the participant time to review the forms independently. Participants were informed about the nature of the study, including potential risks and benefits, and they were informed that they may leave the experiment at any point without penalty. Once participants had reviewed the forms and signed them, the forms were collected. Participants were asked to confirm that they have normal or corrected to normal vision to ensure that they did not have a visual condition that would confound their ability to perform the task.

Practice trials. All participants were seated in front of a computer and desk configuration. The experiment was run within the MIT software, and the researcher recorded performance in the verbal secondary task on a data collection sheet. Participants reviewed the instructions for the task with the researchers, and were allowed to ask questions. A reference sheet included a simplified representation of the secondary task. Participants were also informed they could keep the reference sheet in front of them for the duration of the experiment. The reference sheet was simplified but the altitude and velocity information were flipped from the order typically used to present information (left to right pairs with top to bottom). If participants did not learn the associations, it would impose a higher cognitive workload to have to find the location of altitude and velocity and the correct thresholds because the thresholds were presented in an unusual order. This was done to prevent the reference sheet from becoming a “cheat sheet" that would be easier to read from than to perform the task mentally as intended. The experimental software did not have a pause functionality, so the experiment could not be paused. Therefore, it was necessary to allow participants to keep this information near by in the event they forgot the thresholds or became confused. Referencing the sheet would result in a slower RT, which would be reflected in the performance outcome.

Each participant was required to participate in 15 practice trials to get acquainted with the program (which took between 5-6 minutes). Participants were told if they were not comfortable moving to the experimental stage, they could take the practice trials again. Participants were also asked to repeat the trials if the researcher was concerned they did not understand the task. The data in this stage was not counted toward the participants' performance to remove the risk of the novelty of the experiment negatively impacting performance. These stages were run in the exact same manner as the experimental trials to ensure complete practice on all aspects of the study.

Experimental trials. Once practice trials were complete, the experimental trials began. A total of 100 experimental trials per participant were conducted, broken up into blocks of 50 trials with a 5 minute optional break halfway through. Experimental conditions were counterbalanced between the two blocks. The entirety of the experimental trials took approximately 20 to 30 minutes to complete. Participants were instructed to click “start” once they were ready to begin the trials. Five objects appeared at their designated waypoints within a constraint of at least 1 degree away from the edge of the screen, and at least 1 degree apart upon the start of the trial. No distractor objects were included. All objects on the screen will moved at a consistent speed once the experiment began. The use of distractors in MIT tasks is not needed, because each object's identity is unique and distinct from all other objects (Oksama & Hyönä, 2008).

The program displayed an instruction screen at first. Once participants started the study, a tracking task appeared which was followed by a distraction task (which served as a mask). The participant tracked the stimuli as they moved for 6 seconds. Once 6 seconds passed, the mask screen appeared and occluded the entire tracking task. When the mask lifted, the screen returned to the tracking task and participants identified the location of a target. The program recorded response time, the number of clicks the participant made with the mouse, and the number of times the participant hovered over an object's identity information.

Each object moved around the screen in large, rectangular waypoint paths that were not visible to participants (see Fig. 3). The objects moved along these paths and were often in a position where there was overlap between one or multiple other objects' trajectories which countered the possibility that participants would be able to perform the task by simply memorizing the paths of the objects or the general location of the objects. Using this pattern of movement ensured that objects covered as much distance as possible (displacement) per trial, and moved the exact same distances each time. To ensure this, if a participant reached the end of one length of the rectangle and turned while a mask was present, this object would not be queried. This type of movement also has some external realism (acknowledging that planes do not move in square paths, but move slowly on radar displays in generally straight lines).

A mask was used to occlude a view to the tracked objects in this paradigm. The mask contained an image of a heads up display (HUD) in a jet (Fig. 4). The display contained information about the status of an aircraft (altitude and velocity). Each altitude and velocity value was randomly generated and unique to each trial, meaning 100 combinations of altitudes and velocities were used. Participants were asked to make a determination about whether the number on the left (speed) was above a threshold, or if the other number on the right (altitude) was below a threshold, or "safe". Although participants do not regularly

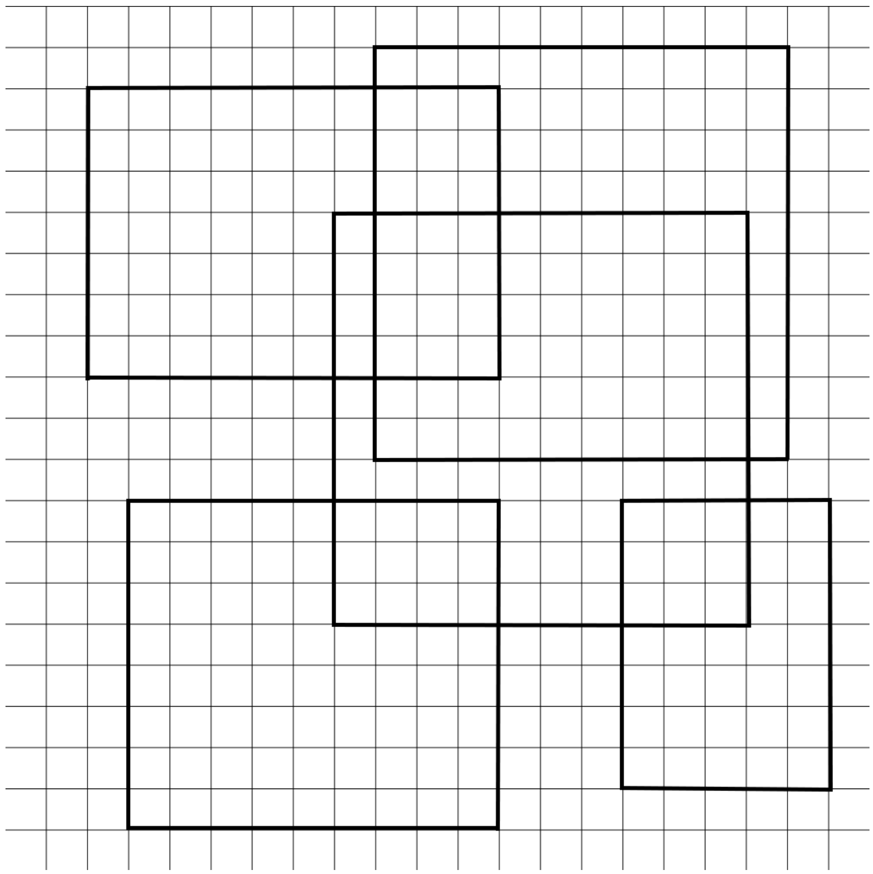


Figure 3. An image of the objects' predetermined trajectories.
Objects moved along these waypoints for the duration of the experiment.

make altitude and velocity judgements, the question was framed as such to help participants grasp and understand the task (participants could think about the task as altitude and velocity judgements rather than determining if one number is above an arbitrary threshold and one number is below an arbitrary threshold) and add face validity. To add further face validity, the altitude judgement was determined to be “safe" if the altitude was above 500. The other judgement is whether velocity was below 1000.

Participants had either 2 or 4 seconds to determine if their altitude was safe (above 500) or the velocity was safe (below 1000). Participants were given the opportunity to practice beforehand to learn the associations.

The secondary task masked the screen for either 2 seconds or 4 seconds. The mask

Figure 4. An image of the mask that features the secondary task.

contained a task that participants attempted to complete before returning to the tracking task. The mask image replicated a HMD. Participants were asked about their "flight status." The display featured a question asking if either the participants' altitude or velocity was in a safe range at the bottom of the screen, to which participants responded "yes" (i.e., in a safe range) or "no" (i.e., not in a safe range) verbally. Each combination of altitudes, velocities, and questions were randomly assigned to each trial, making each trial's mask unique. This served to prevent memorization of flight information as the trials continued. Once the set mask time (i.e., either two or four seconds) had passed, the mask disappeared and the participants will be presented with the tracking screen again.

At this stage of the task, participants were asked to identify the location of a target. The objects froze in place and their labels were covered at the moment the secondary task disappeared. Participants were prompted to locate a target object as quickly as possible and click on it with the mouse. Running the mouse over the occluder box covering an objects label revealed the callsign. Additionally, the objects were no longer moving to prevent errors in the accuracy of the response, and to prevent further object displacement. If participants lost the location-identity bindings, the mask of the object identities served as a way to slow down responding and increase the need to check object identities.

Post-test survey and conclusion. Participants discussed any strategies or techniques they used to keep track of the objects with the researchers. This was used to aid in the identification of any common strategies participants used to perform the task. Then, participants were presented with a debrief and were given an opportunity to ask questions. In total, participants completed the study within 45 to 60 minutes, depending on the speed they took to respond and time to complete the informed consent and conclusion stages.

Data collection. Response time data was collected as a means of measuring performance in the object tracking task. The software logged these data and presented output in a .csv file for analysis. The software also recorded identity check counts and the objects that were clicked on. One object identity check (i.e., verifying the correct object on the first try) was considered to be perfect performance. Trials were removed if the participant clicked on the wrong object as it would not provide useful information about whether or not the participant was able to locate objects. The final count of removed trials will be reported as an indicator of poor performance, but this information cannot be compared to identity check count data (a participant might uncover the label of one object and click on it, but if it is the wrong object, this falsely shows perfect performance). Descriptive and inferential analyses on these performance indicators were performed in R Studio.

Performance in the secondary task via verbal responses was recorded by hand by the researcher on a data collection sheet. Researchers logged either yes, no, or no response (or unclear response). The equation for the serial visual search task was considered as a means of indicating if participants were relying on their memory of location-identity bindings or relying on performing a simple visual search. The following equation was used to identify one's visual search time in serial visual search task (Wickens, Hollands, Banbury, & Parasuraman, 2015):

$$T = NI/2 \tag{5}$$

Where T represents time, N represents number of elements in the visual search field (which will be four in all trials), and I is the average inspection time for each element. This is a measure that was determined to be a poor representation of actual performance in the task based on pilot testing. The difference in the time it took to find the first object was well above the time this equation indicates as the time needed to do a visual search. However, the delay could be attributed to the need to switch tasks before attempting the first try, and the nature of the task between one object reveal and more than one. Once a participant revealed a label, they might have had a general sense of where the object was located, or engaged in a visual search for the object. Therefore, the visual search formula does not account for the different nature of response time between object identity check counts.

Results

Twenty-five participants engaged in a total of 100 trials each. Two participants were removed from analyses; one participant clicked on the incorrect object a total of 61 times (participants were told to search for the correct object before clicking on it), while the other participant responded to the secondary task correctly in only 45% of the trials. Interestingly, both participants performed rather well in the opposite tasks (the participant removed for incorrect object clicks was correct 92% of the time in the secondary task, and the participant that did not perform well on the secondary task had an average of 1.68 identity checks per trial). This suggests participants did not understand the task at hand, or performed poorly in one aspect of the task as a method to compensate for the difficulty of the task and therefore perform better in the other task.

Data from two trials, trials 50 and 100, were removed for all participants. The data for these trials included a mix of unusually long RTs that were not observed or noted by the researchers during the test sessions (the largest being a 738.87 second response time), missing identity check counts ($N = 2$) and missing RTs in these trials. These problems indicate a measurement error in the software for the last trial of each set. It is most likely the case that the timer did not stop running at the conclusion of the trials for some participants. As a result, all of these trials were removed to reduce the risk of measurement error being introduced. Trials 50 and 100 were the last trials in each half of the experiment.

The possibility of clicking on the wrong object by mistake was possible due to the MIT software used to collect data. Once a participant clicked on an object, the trials commenced regardless of which object was clicked on. If participants clicked on the wrong object (either experimental error or by mistake), their response was removed. After removing trials and participants accordingly, a total of 2163 observations remained for analysis in R.

Exploratory Data Analysis

To analyze response time (RT), the RT data were transformed using a natural log due to their skewed nature (Fig. 5). The skew and kurtosis of RT before the transformation was 2.09 and 6.72 in the long occlusion fast movement condition, 2.21 and 10.76 in the long occlusion slow movement condition, 1.35 and 1.92 in the short occlusion fast movement condition, and 1.93 and 5.22 in the short occlusion slow movement condition. Skew and kurtosis were corrected (see Fig. 5 and Fig. 6), bringing the distribution of RT into a normal range.

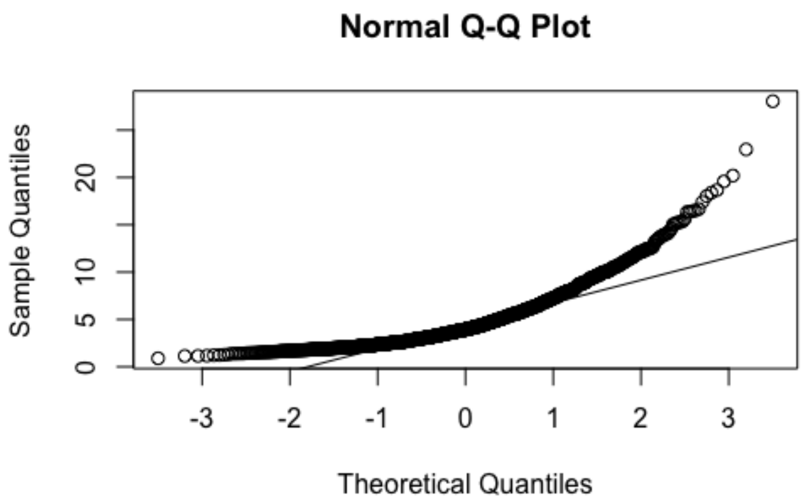


Figure 5. A Q-Q plot of RT values indicates that the data were not normally distributed.

There were no strong indicators of a learning effect taking place based on RT (Fig. 7). The average RT per trial was plotted while there appears to be a slight decrease in performance over time, there is little variance average response time vary very little throughout the experiment. Additionally, there were no indicators that performance was poor based on this variable (Table. 3).

Performance based on identity check counts per condition revealed differences per condition that shows more variation in performance (Fig. 4). Additionally, there was no evidence of a floor or ceiling effect, with participants clicking the correct object on the first try in 59% of the trials and by the second try in 79% of trials. The following tables (Table 4 and Table 5) outlines percentages of each identity check counts per condition and summary

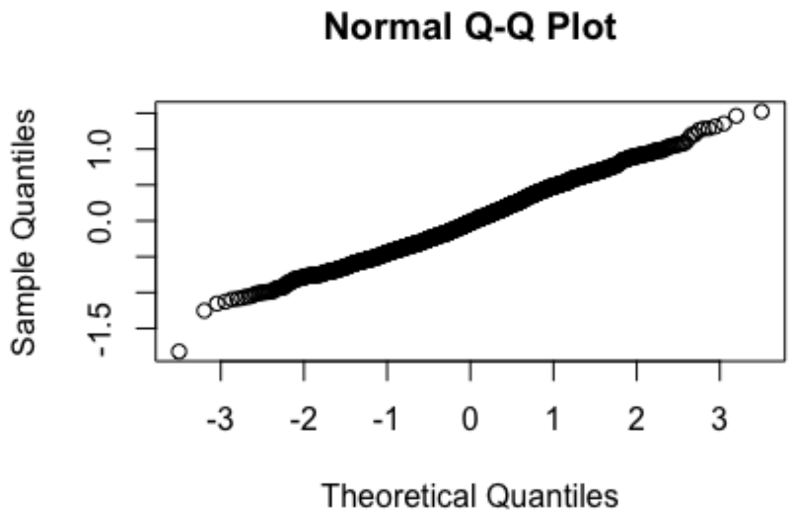


Figure 6. A Q-Q plot of normally distributed RT values after a log transformation.

Table 3. *Summary RT data per condition.*
There is little variance between conditions on the basis of RT.

Condition	Mean	SD	Med	Skew	Kurtosis
Slow Speed, Short Occlusion	1.36	0.53	1.32	0.39	-0.24
Slow Speed, Long Occlusion	1.40	0.52	1.36	0.29	-0.39
Fast Speed, Short Occlusion	1.46	0.56	1.44	0.19	-0.72
Fast Speed, Long Occlusion	1.35	0.52	1.33	0.45	-0.25
Total	1.41	0.53	1.37	0.30	-0.42

statistics by condition.

It was also apparent that there were no learning effects based on identity checks by trial (Fig. 8). The data has an even variance across trials, unlike RT.

Participant performance in the secondary task was measured to obtain the percentage of correct responses. Overall, participants were correct 87% of the time, indicating a strong ability to perform the task. Additionally 15 participants were correct at least 90% of the time. However, performance varied on an individual basis, with four participants being correct less than 75% of the time. One participant was removed after only getting 45% of trials correct, based on being correct less than chance. There were no clear patterns

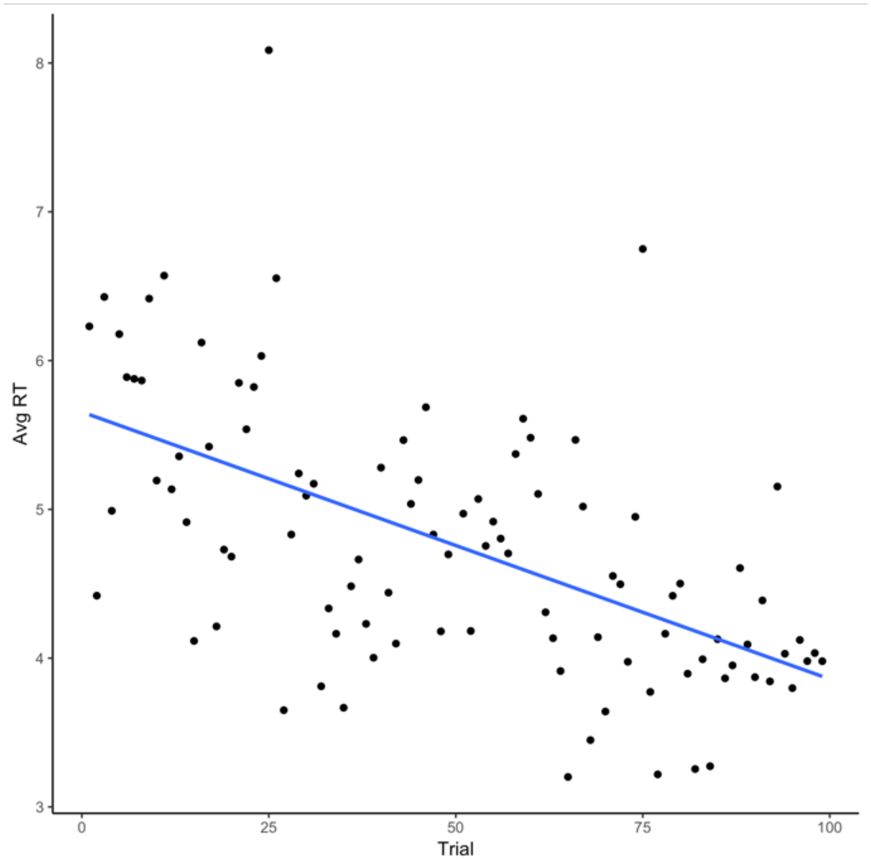


Figure 7. Average RT (s) by trial based on overall performance for all participants. There does not appear to be a clear trend in the data. There is a relatively small range in average transformed RT, with the maximum around 2.

by trial or values presented that indicated why this was the case. Performance in this task was compared to average transformed RT and label reveals (Table 6 and Fig. 9). Pearson correlations indicated that there was no correlation between identity check counts and correct secondary task responses ($R = -0.03$) and a small correlation between correct secondary task responses and RT ($R = 0.25$). Based on this information in the context of all of the results, it can be concluded that performance on the secondary response task was largely independent from performance in the tracking task, but most participants performed well overall.

Table 4. *Percentage of identity checks per condition. Perfect performance was nearly 60% in each condition.*

Condition	One	Two	Three	Four	Five
Slow Speed, Short Occlusion	63%	19%	8%	5%	4%
Slow Speed, Long Occlusion	59%	17%	12%	7%	5%
Fast Speed, Short Occlusion	56%	19%	11%	9%	5%
Fast Speed, Long Occlusion	56%	23%	13%	5%	2%
Total	59%	20%	11%	7%	4%

Table 5. *Summary data of identity check counts per condition. There is little variance between conditions.*

Condition	Mean	SD	Med	Skew	Kurtosis
Slow Speed, Short Occlusion	1.68	1.10	1.00	1.66	1.84
Slow Speed, Long Occlusion	1.82	1.19	1.00	1.33	0.63
Fast Speed, Short Occlusion	1.89	1.22	1.00	1.21	0.28
Fast Speed, Long Occlusion	1.74	1.03	1.00	1.34	1.04
Total	1.78	1.14	1.00	1.39	0.93

RT performance

A plot of interactions between RT using mask duration and speed revealed that there was likely a statistically significant difference between performance in the fast and slow moving object conditions if they had a short occlusion time (Fig. 10). These results have little variance, so while statistical significance may demonstrate differences, the difference in transformed RT is limited to a range of 0.15.

A linear mixed effect model fit by maximum likelihood was conducted using an $\alpha = .05$. A within groups ANOVA was determined to be poorly suited for the analysis of response time due to excluded (missing) RT values. Object speed, mask duration, and order were used as fixed effects, with participant random slopes and intercepts. Based on the results provided in Table 7, order did not appear to have an effect on performance ($p > .05$). Coefficient trends for participant were generally similar, suggesting participants performed about equally well. There was a significant interaction between the slow speed and fast speed condition when the mask duration was short.

There were no violations of normality based on a Q-Q plot (Fig. 6) and a Shapiro-Wilk

Table 6. *Percentage of accuracy per participant.*
The data are paired with average identity check counts and RT to compare general performance.

Correct (%)	Avg IC	Avg transformed RT
99	1.71	1.284
99	1.42	1.303
99	1.66	1.479
99	1.58	1.096
98	2.93	1.409
98	1.12	0.892
97	1.60	1.015
97	2.05	1.567
96	3.13	1.357
96	1.26	1.603
95	2.11	1.192
92	1.02	1.092
92	1.95	1.337
91	1.36	1.526
90	1.43	1.254
87	1.84	1.627
81	2.17	1.711
78	1.12	1.086
77	1.82	1.973
61	1.80	1.797
60	2.35	1.121
60	2.02	1.979
59	1.33	1.333

Table 7. *Significance test results for RT by condition.*
The difference between a short and long mask duration was significant, as well as the interaction between the speed conditions when there is a short mask.

Fixed Effects	Estimate	Std. Error	DF	t	p
Slow speed	0.050	0.027	2140.03	1.837	0.066
Short mask duration	0.116	0.027	2140.042	4.257	$<.001$
Order	-0.066	0.119	22.987	-0.558	0.582
Slow speed, short mask	-0.155	0.038	2140.007	-4.031	$<.001$

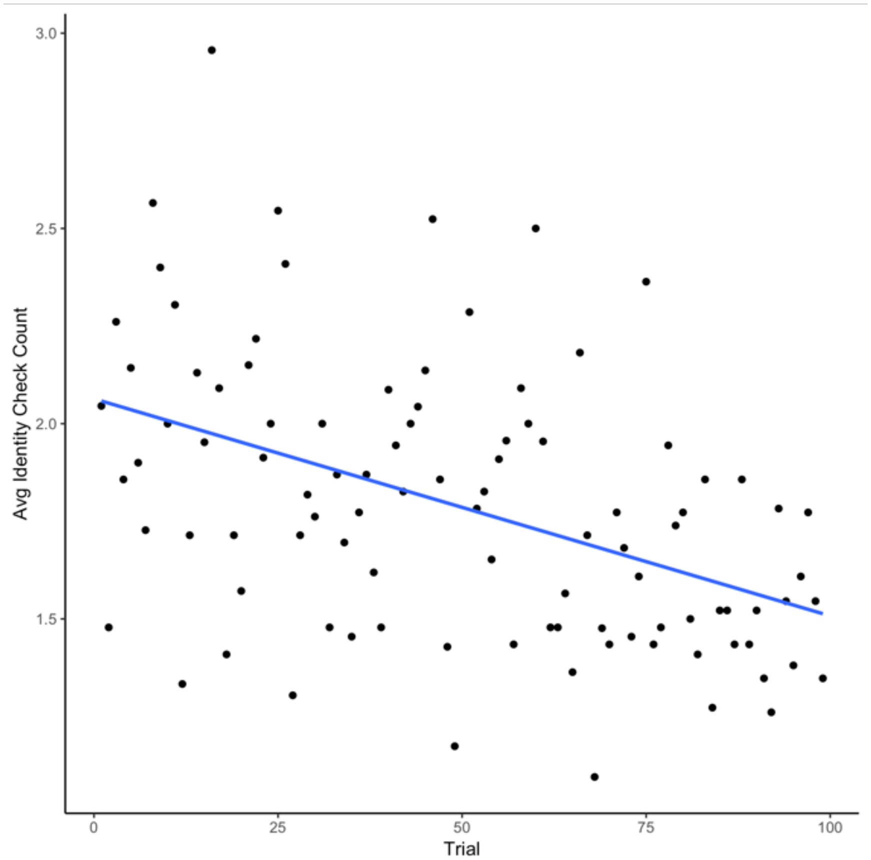


Figure 8. Average number of identity checks per trial.
There is no clear trend in performance based on trial.

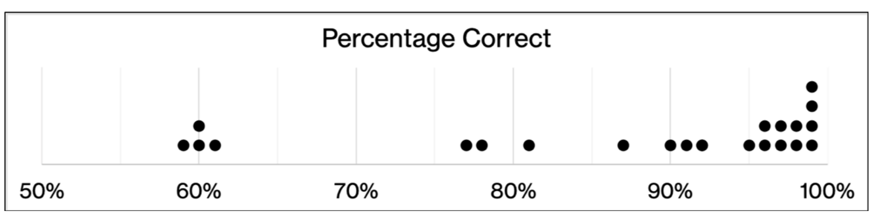


Figure 9. Participants plotted by correct responses.

test ($W = 0.99$, $p = 6.1$). A plot of residuals shows a constant, but small variance (Fig. 11). No violations of multicollinearity were detected ($2 \times DF = 1$ for condition, gender, and order).

Pairwise Tukey comparisons were conducted and adjusted using a Bonferroni method.

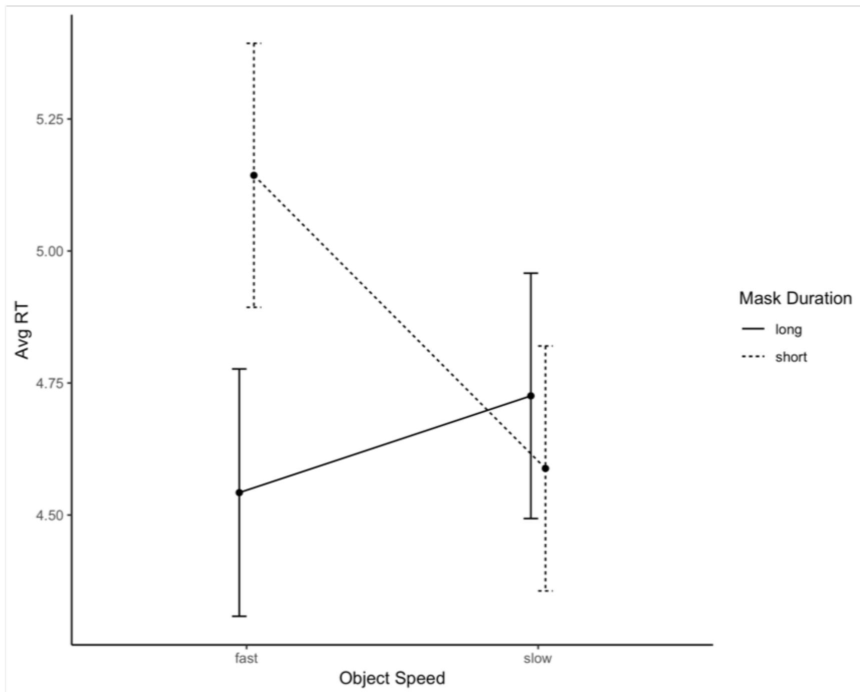


Figure 10. Average RT based on mask duration and speed with 95% CI bars.
It appears that there is a significant interaction between the slow and fast speed conditions when the mask time is short.

There were significant differences in performance between the Fast and Short condition and the Slow and Short condition, as well as the Fast and Short condition (Table 8).

Identity check performance

A plot of interactions between identity check counts using mask duration and speed indicated the potential for a marginal statistically significant differences between performance in the slow and fast conditions during the short masks (Fig. 12). Similarly to RT, these results have little variance, with a difference in range of 0.4.

A chi-square test of goodness-of-fit was run to determine if there were differences in identity check counts per condition. The results indicate that there was a difference in performance between conditions, $\chi^2(2) = 29.205, df = 15, p = .015$. A pairwise comparison of each condition showed three conditions had significant p-values that were negated once the p-values were adjusted p-values using a Bonferroni correction. Therefore it can be

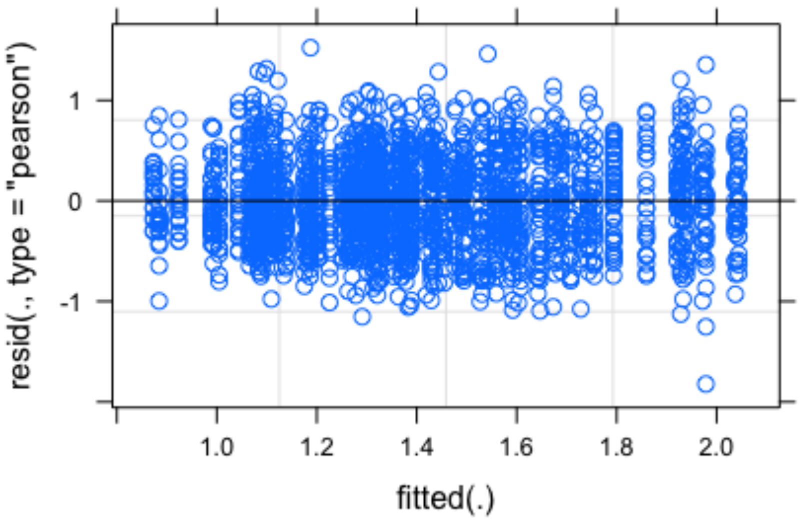


Figure 11. A residual plot showing relatively equal trends in variance.

concluded that identity check counts per condition were not significant. The results are documented below (Table 9 and Table 10).

Self reported strategies

Individuals were asked to identify strategies and techniques used to keep track of the objects upon completing the task upon concluding the experiment. All participants who mentioned the identity checks mentioned that they only used the first letter of the object's

Table 8. *Pairwise comparisons of RT between conditions.*
The comparisons revealed differences between the Fast and Short condition and the Slow and Short condition, as well as the Fast and Short condition.

Comparison	Estimate	Std. Error	z	p
Slow, Long : Fast, Long	0.051	0.027	1.837	0.397
Fast, Short : Fast, Long	0.116	0.027	4.257	<0.001
Slow, Short : Fast, Long	0.011	0.027	0.418	1.000
Fast, Short : Slow, Long	0.065	0.027	2.414	0.094
Slow, Short : Slow, Long	-0.040	0.027	-1.437	0.903
Slow, Short : Fast, Short	-0.105	0.027	-3.881	<0.001

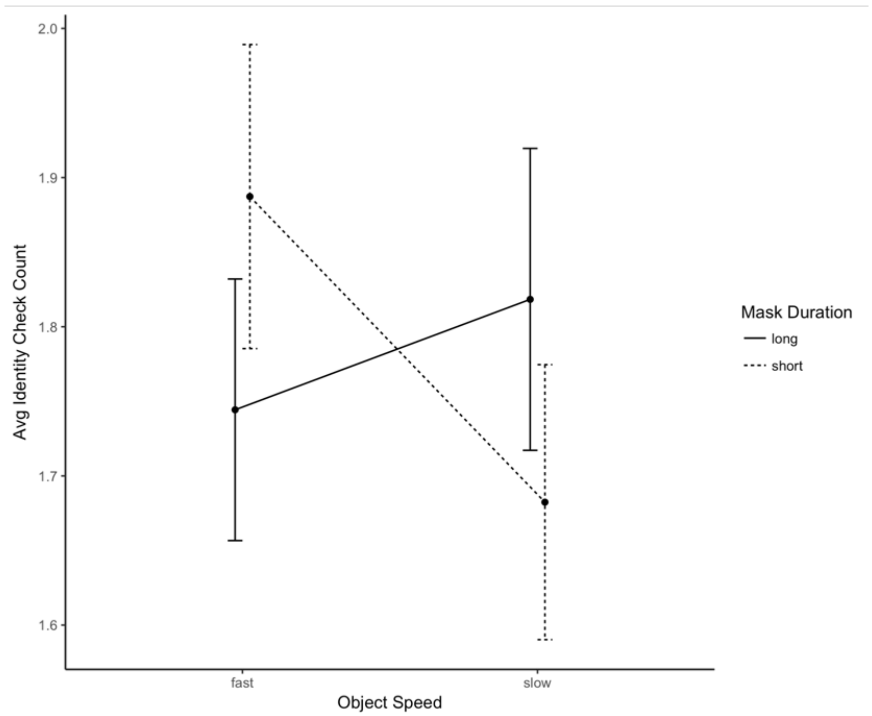


Figure 12. Average identity check counts by mask duration and speed with 95% CI bars. There appears to be a small difference in performance between the slow and fast conditions when the mask duration was short.

identity (N = 15). Several participants used acronyms as the objects would move (such as "MTF" when the objects were in this order). Many participants tried to divide the screen into quadrants or sections and check each section to remember which objects were in the section. These strategies might explain why participants were not successful on the first try, but they were by the second try. Other participants tried to read the order of the objects from top to bottom, or left to right, or in a clockwise order. Two participants noted that they did not feel that they were trying to keep track of the objects, but they were successfully guessing. It is unlikely this were the case, but it could suggest that they did not need to expend as much attentional resources as other individuals to keep track of the objects. One participant noted that they actually only tracked three objects, and guessed if the other two were queried.

Table 9. *Significance test results for identity check counts by condition.*
There were no statistically significant differences between condition after a Bonferroni correction was applied.

Condition	p.Chisq	p.adj.Chisq
Long Fast : Long Slow	0.015	0.091
Long Fast : Short Fast	0.012	0.070
Long Fast : Short Slow	0.009	0.057
Long Slow : Short Fast	0.679	1.000
Long Slow : Short Slow	0.178	1.000
Short Fast : Short Slow	0.030	0.182

Table 10. *A contingency table of identity check counts.*

Condition	1(n)	1(%)	2(n)	2(%)	3(n)	3(%)	4(n)	4(%)	5(n)	5(%)
Long Fast	297	56%	122	23%	68	13%	29	5%	12	2%
Long Slow	316	59%	91	17%	62	12%	38	7%	27	5%
Short Fast	306	56%	106	19%	61	11%	48	9%	29	5%
Short Slow	349	63%	104	19%	46	8%	28	5%	24	4%

Task switching cost

An apparent trend in the data warranted further analysis: the amount of time it took to respond and identify the first object was longer than it took to find subsequent objects. Each participant's average RT can be seen compared to the average time it took to identify subsequent objects if the object was not identified on the first time, otherwise known as the visual search time (Table 11). Participants' average visual search times were between 2.24 and 5.09 seconds less than response times, with an average difference of 3.83 seconds.

Table 11. *Average RT and visual search time.*
The data are followed by the difference between the two values by participant.

Participant	Average RT	Average Visual Search	Time Difference
1	4.744	1.047	3.697
2	4.786	1.226	3.560
3	4.595	1.083	3.511
4	5.409	1.422	3.987
5	4.795	1.650	3.145
6	5.643	1.503	4.140
7	3.699	0.753	2.946
8	5.950	1.575	4.375
9	5.609	1.296	4.312
10	5.955	1.277	4.679
11	6.202	1.107	5.095
12	3.193	0.954	2.240
13	3.435	1.120	2.315
14	5.012	1.148	3.864
15	5.251	1.172	4.080
16	4.532	1.011	3.520
17	4.117	0.914	3.202
18	4.759	1.089	3.670
19	6.247	2.325	3.922
20	5.506	0.915	4.591
21	6.497	1.538	4.959
22	5.799	2.022	3.777
23	5.422	0.828	4.594

Discussion

The present study has been an analysis of sources of poor performance in MIT tasks. Specifically, two problems of interest were addressed; the first being does the duration of an occlusion or the displacement that occurs during an occlusion have a greater impact on performance in an MIT task. Secondly, a secondary task was added to prevent maintenance rehearsal of objects during the occlusion, which is representative of how objects are tracked through occlusions in realistic scenarios. By knowing which aspects of tracking contribute to poor performance, measures can be taken to reduce the likelihood of these scenarios from occurring in operational settings, or designs can be adjusted to reduce one's cognitive workload. Considerations of the findings related to performing secondary tasks could be applied to the design of products and systems in the event operators are in a position where they are handling multiple tasks at once.

Participants were asked to track multiple moving objects through either a 2 or 4 second mask, and perform a secondary task when the moving objects were occluded from view. Participants were relatively successful at performing both tasks, although several participants struggled to perform the secondary masking task (with performance below 75%). Individual differences have been known to impact tracking performance historically (Oksama & Hyönä, 2008). This appears to be the case in this scenario, as several participants could not perform a secondary task nearly as well as others while tracking objects in this study. The specific masks that individuals struggled to respond to did not have clear patterns or obvious reasons as to why specific masks could have been more difficult than others. For example, a velocity of 400 would be a safe velocity but not a safe altitude, but there was not a pattern of responding to these more challenging masks incorrectly. The secondary task was designed to keep participants occupied for several seconds and participants, yet 14 participants were able to perform at 90% accuracy, and four were able to perform at 99% accuracy. This further supports the case for individual differences playing a role in the outcome of this task, and it provides strong evidence of the ability of participants to track multiple moving objects at once and perform a secondary task.

Response time in the present study yielded statistically significant differences between the speed of the objects during a fast and short (2 second) mask compared to a fast and long mask, and a slow and short mask. The practical difference in RT should be taken into account; average RTs were particularly long for a tracking study (over 4.5 seconds) and the difference was in a range of .075 seconds. The lack of significance when the mask was longer cannot be explained by these results. It is possible that the task became more difficult regardless of displacement. Yet, the mean RT per condition did not get longer (and therefore worse) as a function of mask length. It is likely that response time was not a useful measure of performance in this task regardless of significance, and RT did not capture performance in the way it has in previous studies on MIT. In particular, RT may have been similar no matter how well the participant was able to track objects because participants could have taken their time to identify the correct object on the first try, and then engaged in a quick visual search if the first attempt was incorrect.

The use of the visual search equation in this study revealed that there was a cost to switch between the secondary and tracking task. Typically, when participants are engaged in a visual search task, it takes a given amount of time to search the visual field and identify the correct object. Comparing participants' RTs to the average visual search time would help reveal which participants were engaged in the tracking task, and which participants were simply performing a visual search. However, the results we found are unintuitive because participants performed well in other standards, but took an average of 3.834 seconds longer to find the first object compared to the subsequent objects. There are several possible reasons for this. The first is that participants are suffering from a delayed RT from switching between the secondary task and back to the secondary task. Participants having a delayed RT is consistent with previous studies on task switching literature (Wylie & Allport, 2000). The second is that participants might be using strategies to identify the target object on the screen, and need time to think through his or her chosen strategy. Once the participant realizes an object is not the correct object, the participant likely had a second object in mind, or they needed to engage in a visual search. The time it would take to find the correct object without trying to recall the correct object is intuitively less than simply searching for the object. We can conclude the visual search equation is not likely an inappropriate measure of performance. Rather, identity check counts provide more useful information about participant performance.

Identity check count data indicated that participants could correctly identify the correct object in 59% of the trials and by the second object reveal in 79% of trials. Revealing objects correctly on the second try is likely an indicator that participants maintained the relative location of an object, but could have confused objects that were in close proximity together. It is worth considering the likelihood that individuals can maintain the relative location of objects, but might not extrapolate motion and be able to track objects perfectly through occlusions. The nature of how the objects moved on the screen meant that they were often somewhat close together. The percentages of performance suggest neither a floor or ceiling effect, but it is poorer performance compared to what has been found historically. In a study by Keane and Pylyshyn (2006), participants could successfully track objects through 900 ms occlusions in 90-80% of the trials, and there was almost no difference between the 900 ms and shorter occlusions. It is likely that the task switching task was the main detriment to performance. The relatively similar RT and identity check count values between conditions lends evidence to this. Additionally, the lack of statistical significance between all conditions for object reveal counts and RT supports this. If there were an aspect of the tracking task that contributed to poor performance, it would have been more apparent from statistical testing. It is possible that the differences between displacements in conditions were not large enough to impact performance, because the duration of the mask conditions did not impact performance.

The findings support the ability of individuals to track multiple moving objects and perform a secondary task with relative accuracy, but not without difficulty. A rate of 59% performance would not be considered acceptable in an operational setting. Additionally, performance between the secondary task and RT or identity check counts was not correlated, meaning performance in the tracking task was largely unrelated to performance in the secondary task. Yet, participants did very well in the secondary task overall. Additionally, it is possible to track multiple moving objects, maintain their object identities, and perform a secondary task. There are several possible explanations for why this is. The study provides evidence that the mechanisms that maintain the memory of locations and identities in object tracking are likely separate from the mechanisms that allow individuals to perform other tasks, or else performance would not have been possible at all. It is possible that giving participants six seconds to track objects was enough time to encode the relative location of objects without needing maintenance rehearsal to maintain those bindings. It is also possible that participants could quickly recover the identity bindings of objects once they were able to view the locations of objects (the labels were covered, but the actual location of objects was visible). This study provides evidence of a strong memory mechanism for identity-location bindings that can endure for several seconds.

The self reported strategies were insightful as to how participants performed the task. The division of the screen into sections or grouping objects might explain why participants were not successful on the first try, but they were by the second try. Participants could have used these strategies to maintain the relative location of objects, but not the exact position. Trimming the object identity names to the most simple version possible is useful when operators must keep track of objects with lengthy identities, but this could be costly when objects have similar names. This information suggests that participants are not extrapolating motion, but using memory techniques to try and encode information. It is also insightful to know that participants might not know the object on the first try alone, but using strategies can help individuals group objects. This is worth further investigation, and could be useful for individuals in operational settings to strategize when performing these tasks.

Conclusion

The ability to track multiple moving objects and their identities while switching between a secondary task is possible, and a secondary task can be maintained while doing so. However, these findings do not provide definitive conclusions about motion extrapolation. Furthermore, engaging in a secondary task during an MIT task can be somewhat detrimental to performance and should be carefully managed in operational settings. The occlusion of objects over several seconds does not impact performance in MIT, which is supported by previous studies on MOT and a limited amount of MIT research. The results suggest that one's response time is slower after a performing a secondary task, so care should be taken to prevent individuals from having to perform secondary tasks as much as possible in operational settings when the maintenance of object location-identity bindings is paramount to safe performance in said setting. These findings also support the use of mnemonic devices and other devices to aid with the encoding of location-identity bindings. These can be trained or suggested to individuals performing these tasks in operational settings.

Recommendations

Participants could accidentally click on the wrong object in the present study, which we would consider to be a limitation. It is likely that a number of instances where data was removed could have resulted from participants clicking on an object by mistake when they meant to simply check the label, or incidences where participants thought they knew the location of the object and clicked on the wrong one prematurely. Therefore, we recommend using programs that do not allow participants to advance to the next trial unless participants click on the correct object.

The number of novel factors in this study require further evaluation. Comparing performance in tracking tasks with and without a secondary task would further clarify how a secondary tracking task impacts tracking performance, while also providing further clarification on the differences in tracking task conditions. Further study on how long individuals can maintain location-identity bindings would be essential for determining the types of secondary tasks that can be performed while individuals are tracking objects. It is worth examining these relationships and comparing them to other factors known to influence tracking performance, such as the number of objects in the tracking task and object entropy.

References

Alvarez, G. A., Horowitz, T. S., Arsenio, H. C., DiMase, J. S., & Wolfe, J. M. (2005). Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 643–667.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, *4*(11), 417–423.

Clair, R. S., Huff, M., & Seiffert, A. E. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision*, *10*, 1–13.

Cohen, M. A., Pinto, Y., Howe, P. D. L., & Horowitz, T. S. (2011). The what-where trade-off in multiple-identity tracking. *Attention, Perception, and Psychophysics*, *73*, 1422-1434.

Endsley, M. R. (1987). Sagat: A methodology for the measurement of situation awareness (nor doc 87-83). *Hawthorne, CA: Northrop Corporation.*

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*, 65–84.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 32–64.

Endsley, M. R., & Garland, D. J. (2000). *Situation awareness analysis and measurement.* CRC Press.

Fencsik, D. E., Klieger, S. B., & Horowitz, T. S. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception Psychophysics*, *69*, 567–577.

Franconeri, S. L., Pylyshyn, Z. W., & Scholl, B. J. (2012). A simple proximity heuristic allows tracking of multiple objects through occlusion. *Attention, Perception, Psychophysics*, *74*, 691–702.

Gertler, J. (2012). F-35 joint strike fighter (jsf) program..

Hope, R. M., Rantanen, E. M., & Oksama, L. (2010). Multiple identity tracking and entropy in an atc-like task. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 54, pp. 1012–1016).

Horowitz, T. S., Birnkrant, R. S., Fencsik, D. E., Tran, L., & Wolfe, J. M. (2006). How do we track invisible objects? *Psychonomic Bulletin Review*, *13*, 516–523.

Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception Psychophysics*, *69*, 172–184.

Howard, C. J., Masom, D., & Holcombe, A. O. (2011). Position representations lag behind targets in multiple object tracking. *Vision Research*, *51*, 1907–1919.

Howe, P. D., & Holcombe, A. O. (2012). Motion information is sometimes used as an aid to the visual tracking of objects. *Journal of vision*, *12*, 1–10.

Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? tracking as a low-level, non-predictive function. *Cognitive Psychology*, *52*, 346–368.

Luu, T., & Howe, P. D. (2015). Extrapolation occurs in multiple object tracking when eye movements are controlled. *Attention, Perception, Psychophysics*, *77*, 1919–1929.

Meyerhoff, H. S., Papenmeier, F., & Huff, M. (2017). Studying visual attention using the multiple object tracking paradigm: A tutorial review. *Attention, Perception, & Psychophysics*, *79*(5), 1255–1274.

Moray, N. (1984). Attention to dynamic visual displays in man-machine systems. *Varieties of attention*, 485–513.

Nalbandian, A., & Rantanen, E. (2015). Learning of location-identity bindings: Development of level 1 situation awareness in an air traffic control-like task. In *Proceedings of the 18th international symposium on aviation psychology* (pp. 159–164).

Oinonen, K., Oksama, L., Rantanen, E., & Hyönä, J. (2009). Do velocity vectors support multiple object tracking? In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 53, pp. 16–20).

Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? an individual difference approach. *Visual cognition*, *11*, 631–671.

Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive psychology*, *56*, 237–283.

Oksama, L., & Hyönä, J. (2016). Position tracking and identity tracking are separate systems: Evidence from eye movements. *Cognition*, *146*, 393–409.

Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking: I. tracking without keeping track of object identities. *Visual cognition*, *11*, 801–822.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, *3*, 179–197.

Saha, P. K., Samanta, D., Sarcar, S., & Sharma, M. K. (2012). Analysis of visual search features. *International Journal of Human Factors Modelling and Simulation*, *3*(1), 66–89.

Shooner, C., Tripathy, S. P., Bedell, H. E., & Öğmen, H. (2010). High-capacity, transient retention of direction-of-motion information for multiple moving objects. *Journal of Vision*, *10*, 1–20.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology & human performance.* Psychology Press.

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs". *Psychological research*, *63*(3-4), 212–233.

Zhong, S.-h., Ma, Z., Wilson, C., Liu, Y., & Flombaum, J. I. (2014). Why do people appear not to extrapolate trajectories during multiple object tracking? a computational investigation. *Journal of Vision*, *14*, 1–30.

Appendix

Appendix A

R Code:

```
# Install packages
if(!require(psych)){install.packages("psych")} # Summary statistics
if(!require(rcompanion)){install.packages("rcompanion")} # Log transformation
if(!require(MASS)){install.packages("MASS")} # Chi-square calculation
if(!require(ggplot2)){install.packages("ggplot2")} # Plotting
if(!require(Rmisc)){install.packages("Rmisc")} # Plotting
if(!require(Hmisc)){install.packages("Hmisc")} # Plotting
if(!require(ggpmisc)){install.packages("ggpmisc")} # Plotting
if(!require(gridExtra)){install.packages("gridExtra")} # Plotting
if(!require(lme4)){install.packages("lme4")} # LME
if(!require(lmerTest)){install.packages("lmerTest")} # LME
if(!require(multcomp)){install.packages("multcomp")} # Post hoc
if(!require(car)){install.packages("car")} # Post hoc

# Clear preexisting variables in RStudio
rm(list = ls())

# Import dataset
data.mit = read.csv("combined_data.csv", header = TRUE)

# Identify missing data points
which(is.na(data.mit$rt))

# Omit missing data points (NA's)
data.mit <- na.omit(data.mit)
```

```
# Log transform RT
rt_log = log(data.mit$rt)
library(rcompanion)
data.mit$rt_log = rt_log # Add transformed RT to data frame

# Summary statistics
str(data.mit)
library(psych)
describeBy(data.mit,data.mit$condition)

# LME for RT
library(lmerTest)
library(lme4)
rt.null = lmer(rt_log ~ speed + maskdur + order + (1+speed|participant) +
(1+maskdur|participant), data=data.mit, REML=FALSE)
rt.model = lmer(rt_log ~ speed*maskdur + order + (1+speed|participant) +
(1+maskdur|participant), data=data.mit,REML=FALSE)
rt.anova = anova(rt.null,rt.model)
summary(rt.model)

reduced.model = lmer(rt_log ~ condition + order + (1|participant),
data=data.mit,REML=FALSE)
posthoc <- glht(reduced.model, linfct = mcp(condition = "Tukey"))
summary(posthoc, test = adjusted("bonferroni"))

# Assumptions for LME
library(car)
```

```
plot(rt.model,add.smooth=F) # Constant variance of residuals
shapiro.test(residuals(rt.model)) # Normal distribution
qqnorm(residuals(rt.model)) # Q-Q plot for normality
vif(rt.model) #colinearity (variance inflation factors)

# Chi-square statistic for identity checks
library(MASS)
library(rcompanion)
library(multcomp)
object_reveal = table(data.mit$condition, data.mit$hovercount)
object_reveal
chisq.test(object_reveal)
pairwiseNominalIndependence(object_reveal, fisher =FALSE,
gtest=FALSE,chisq=TRUE, method="bonf")

# Correlation between verbal, RT, and identity check data
# Export to excel to get rt_log averages
write.csv(data.mit, file = "rt_log.csv")

# Create matrix with average values to correlate
verbal = matrix(c(1.72,2.18,1.63,1.44,2.05,1.80,2.37,1.36,2.02,1.16,1.04,3.13,
2.92,1.85,1.68,1.93,2.10,1.57,1.84,1.16,1.27,1.36,1.43,99,81,97,99,
97,61,60,91,60,78,92,96,98,87,99,92,95,99,77,98,96,59,90,1.284,
1.711,1.015,1.303,1.567,1.797,1.121,1.526,1.979,1.086,1.092,1.357,
1.409,1.627,1.479,1.337,1.192,1.096,1.973,0.892,1.603,1.333,1.254),
                nrow=23,ncol=3)
x_hc <- verbal[ ,1]
y_hc <- verbal[ ,2]
```

```
hc_cor <- cor(x_hc, y_hc,method = c("pearson"))
res_hc <- cor.test(x_hc, y_hc,method = c("pearson"))
res_hc

x_rt <- verbal[ ,1]
y_rt <- verbal[ ,3]
rt_cor <- cor(x_rt, y_rt,method = c("pearson"))
res_rt <- cor.test(x_rt, y_rt,method = c("pearson"))
res_rt

## Overview plots
library(ggplot2)
library(Rmisc)
library(Hmisc)
library(ggpmisc)
library(gridExtra)

# RT normality before log transformation
qqnorm(data.mit$rt)
qqline(data.mit$rt)

# box plot for LME
boxplot(rt_log ~ speed*maskdur,
col=c("white","lightgray"),data.mit)

# Data frame for trial plotting
data.trial <- data.mit[,c(2,4,5,9,10,11)]
data.trial$order <- as.factor(data.trial$order)
```

```
# RT by trial
# Linear regression of RT
fit.rt <- lm(rt ~ trial2, data = data.trial)
summary(fit.rt)

# Plot RT
theme_set(theme_classic())
trial_rt_summary = summarySE(data.trial, measurevar="rt",groupvars=c("trial2"))
plot1 <- ggplot(trial_rt_summary, aes(x=trial2, y=rt)) +
geom_point() +
geom_smooth(method = "lm", se=FALSE) +
xlab("Trial") + ylab("Avg RT")
ggsave("rt_trial.png")

# Identity checks by trial

# Linear regression of identity checks
fit.hc <- lm(hovercount ~ trial2, data = data.trial)
summary(fit.hc)

# Plot identity checks
theme_set(theme_classic())
trial_hc_summary = summarySE(data.trial, measurevar="hovercount",groupvars=c("trial2"))
plothc1 <- ggplot(trial_hc_summary, aes(x=trial2, y=hovercount)) +
geom_point() +
geom_smooth(method = "lm", se=FALSE) +
xlab("Trial") + ylab("Avg Identity Check Count") +
```

```
ggsave("hc_trial.png")

# Speed x mask duration interactions - identity checks
hc_summary <- summarySE(data.mit, measurevar="hovercount",groupvars=c("speed","maskdur"))
pd <- position_dodge(0.1) # move .05 to the left and right
ggplot(hc_summary, aes(x=speed, y=hovercount, linetype=maskdur, group=maskdur)) +
geom_errorbar(aes(ymin=hovercount-ci, ymax=hovercount+ci), width=.1, position=pd) +
geom_line(position=pd) +
geom_point(position=pd) +
xlab("Object Speed") + ylab("Avg Identity Check Count") +
labs(linetype="Mask Duration") +
ggsave("hc_int.png")

# Speed x mask duration interactions - RT
rt_summary = summarySE(data.mit, measurevar="rt",groupvars=c("speed","maskdur"))
pd <- position_dodge(0.1) # move .05 to the left and right
ggplot(rt_summary, aes(x=speed, y=rt, linetype=maskdur, group=maskdur)) +
  geom_errorbar(aes(ymin=rt-ci, ymax=rt+ci), width=.1, position=pd) +
  geom_line(position=pd) +
  geom_point(position=pd) +
  xlab("Object Speed") + ylab("Avg RT") + labs(linetype="Mask Duration")
ggsave("rt_int.png")
```

Appendix B

R·I·T

Rochester Institute of Technology

RIT Institutional Review Board for the
Protection of Human Subjects in Research
141 Lomb Memorial Drive
Rochester, New York 14623-5604
Phone: 585-475-7673
Fax: 585-475-7990
Email: hmfsrs@rit.edu

Form C
IRB Decision Form
FWA# 00000731

TO: Ashley Buck

FROM: RIT Institutional Review Board

DATE: **January 17, 2019**

RE: Decision of the RIT Institutional Review Board

Project Title – Multiple Identity Tracking and Motion Extrapolation

The Institutional Review Board (IRB) has taken the following action on your project named above.

☒ Exempt _46.101 (b) (2)_

Now that your project is approved, you may proceed as you described in the Form A.

You are required to submit to the IRB any:

- **Proposed** modifications and wait for approval before implementing them,
- Unanticipated risks, and
- Actual injury to human subjects.

Appendix C

Multiple Identity Tracking and Motion Extrapolation

Informed Consent Form

INTRODUCTION
You are invited to join a research study to look at how people are able to track multiple moving objects at once. Please take whatever time you need to discuss the study with your family and friends, or anyone else you wish to. The decision to join, or not to join, is up to you.
In this research study, we are investigating how people are able to track multiple moving objects through "occlusions," which means the objects are blocked from view, and while performing a second task.

WHAT IS INVOLVED IN THE STUDY?
If you decide to participate you will be asked to complete a series of object tracking tasks. You will be playing the role of a pilot flying a jet. There will be other "jets" on the screen, represented by small jet icons with a series of numbers and letters next to them that you will be in charge of keeping track of. You will need to try to remember where the objects are located on the screen, and the name of each flight. At random times during the task, an image of a display will appear with information related to your "flight." You will be tasked with quickly checking if your altitude and velocity are in a safe range. You will left click the mouse if you determine that the numbers are in a safe range, and you will right click the mouse if the numbers are not in a safe range. After performing this task, you will be asked to locate a specific flight as quickly as you can. This will repeat many times.

When you are done with these tasks, we will briefly ask you for some information about yourself. This will include you race, gender, and whether or not you play video games. We think this will take you 60 minutes.

The investigators may stop the study or take you out of the study at any time they judge it is in your best interest. They may also remove you from the study for various other reasons. They can do this without your consent.

You can stop participating at any time. If you stop you will not lose any benefits.

RISKS
During this study, you might experience feelings of frustration or stress from the difficulty of the task. There may also be other risks that we cannot predict.

BENEFITS TO TAKING PART IN THE STUDY?
By participating in this study, you might learn about how Psychology studies, particularly Perceptual studies work. You might also find the experience to be a fun or enjoyable challenge. Others may benefit in the future from the information we find in this study.

CONFIDENTALITY
We will take the following steps to keep information about you confidential, and to protect it from unauthorized disclosure, tampering, or damage: your personal information will be tied to a pin number and referred to only as that number in all forms other than your informed consent form. Your informed consent form will be kept in a locked file that will only be accessible by the primary investigator and faculty supervisor if needed. No information will be tied to your name in any written outcomes of the study.

INCENTIVES
You will be compensated with SONA system credits, or entry into a raffle to win a $50 Amazon gift card.

YOUR RIGHTS AS A RESEARCH PARTICIPANT?
Participation in this study is voluntary. You have the right not to participate at all or to leave the study at any time. Deciding not to participate or choosing to leave the study will not result in any penalty or loss of benefits to which you are entitled, and it will not harm your relationship with the researchers, or your professor if you are taking part in this study for course credit.
To withdraw from the study, please say so to the researcher in the room with you, and you will be allowed to leave. If you are comfortable, we will ask you why you decided to leave, or follow-up through an email if you would rather not answer in person. Responding to this request is not mandatory in any way, it is only meant to inform the researchers about your decision and assess if your reason for leaving would impact other participants after you.

CONTACTS FOR QUESTIONS OR PROBLEMS?
Call Ashley Buck at (315)877-6710 or email at aeb1530@rit.edu if you have questions about the study, any problems, unexpected physical or psychological discomforts, any injuries, or think that something unusual or unexpected is happening.

You may also contact Dr. Esa Rantanen at (585) 475-4412 or email at esa.rantanen@rit.edu. Contact Heather Foti, Associate Director of the HSRO at (585) 475-7673 or hmfsrs@rit.edu if you have any questions or concerns about your rights as a research participant.

Consent of Subject (or Legally Authorized Representative)
Signature of Subject or Representative Date

____

Appendix D

Debriefing Statement

Dear Participant:

Thank you for participating in this study.

This study examined the theory of "Multiple Identity Tracking." Specifically, this study examined whether participants would be able to track multiple moving objects and perform a secondary task while doing so. Additionally, our research goal is to determine whether any difficulties you might have experienced were due to how far the objects moved while you were unable to see them, or how long you were unable to see them. We expected that the distance the object moved had a greater influence on your performance.

You might have experienced some difficulty performing the task today. This task was designed to be challenging and does not reflect on your intelligence or capabilities.

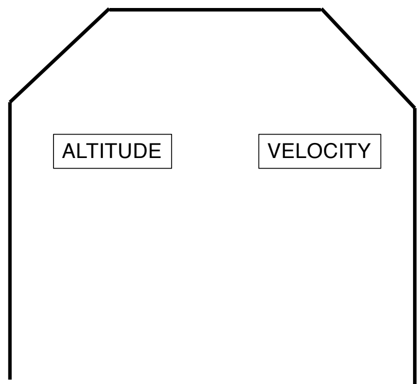
We would like to remind you that this is an ongoing study that will continue to run. For this reason, we ask you not to discuss this experiment with anybody except the researcher.

Once again, thank you for your participation. If you have any questions regarding this study, feel free to contact Ashley Buck at aeb1530@rit.edu or the Human Subjects Research Office at Rochester Institute of Technology (585) 475-7673.

Appendix E

Reference Sheet

REFERENCE SHEET



SAFE VELOCITY = BELOW 1000

SAFE ALTITUDE = ABOVE 500

Appendix F

Task Instructions

INSTRUCTIONS

In this study, you are a the 'pilot' of a jet. You will be asked to keep track of moving objects with labels on a screen. Your job is to keep track of where the objects are and remember them by their label.

Every few seconds, a screen will pop up that looks like your reference sheet. You will be asked to identify if you are at a safe altitude or velocity.

Velocity: A safe velocity is below 1000.

- **Safe**: If you are flying **below** the 1000, say "**yes**" out loud.
- **Unsafe:** If you are flying **above** 1000, say "**no**" out loud.

Altitude: A safe altitude is above 500.

- **Safe**: If you are flying **above** 500, say "**yes**" out loud.
- **Unsafe:** If you are flying **below** 500, say "**no**" out loud.

The location of your altitude and velocity are located on the reference sheet. You can use your reference and instruction sheets at any time.

Once this screen goes away, you will need to find the location of a specific object. When the text saying "Click on Flight <label>" appears, click the correct object.

The labels will be covered, but you can hover over the label to reveal the text. Revealing the text will be recorded as a part of your performance.